Interactive music:

Balancing creative freedom with musical development

Tim Murray-Browne

Thesis submitted in partial fulfilment of the requirements of the University of London for the Degree of Doctor of Philosophy

Centre for Digital Music
School of Electronic Engineering and Computer Science
Queen Mary University of London

Submitted: April 2012
Revised: October 2012

I certify that this thesis, and the research to which it refers, are the product of my own work, and that any ideas or quotations from the work of other people, published or otherwise, are fully acknowledged in accordance with the standard referencing practices of the discipline. I acknowledge the helpful guidance and support of my supervisors, Prof Mark D. Plumbley and Dr Nick Bryan-Kinns.

Abstract

This thesis is about *interactive music*—a musical experience that involves participation from the listener but is itself a composed piece of music—and the Interactive Music Systems (IMSs) that create these experiences, such as a sound installation that responds to the movements of its audience. Some IMSs are brief marvels commanding only a few seconds of attention. Others engage those who participate for considerably longer. Our goal here is to understand why this difference arises and how we may then apply this understanding to create better interactive music experiences.

I present a refined perspective of interactive music as an exploration into the relationship between action and sound. Reasoning about IMSs in terms of how they are subjectively perceived by a participant, I argue that fundamental to creating a captivating interactive music is the evolving cognitive process of making sense of a system through interaction.

I present two new theoretical tools that provide complementary contributions to our understanding of this process. The first, the *Emerging Structures* model, analyses how a participant's evolving understanding of a system's behaviour engages and motivates continued involvement. The second, a framework of *Perceived Agency*, refines the notion of 'creative control' to provide a better understanding of how the norms of music establish expectations of how skill will be demonstrated.

I develop and test these tools through three practical projects: a wearable musical instrument for dancers created in collaboration with an artist, a controlled user study investigating the effects of constraining the functionality of a screen-based IMS, and an interactive sound installation that may only be explored through coordinated movement with another participant. This final work is evaluated formally through discourse analysis.

Finally, I show how these tools may inform our understanding of an oft-cited goal within the field: conversational interaction with an interactive music system.

Acknowledgements

I am enormously thankful and indebted to my supervisors Prof Mark D. Plumbley and Dr Nick Bryan-Kinns for their tireless advice and support throughout this PhD.

I acknowledge with considerable gratitude the contribution of Di Mainstone in the design, construction and subsequent performances of the Serendiptichord of Section 3.2, of Prof Rosemary Bailey to the experimental design of Section 5.3.2 and the statistical analysis of Section 5.5.1, and of Tiff Chan who provided creative input into the development of Impossible Alone of Section 7.3.

Many thanks to those who read draft chapters and papers along the way: my supervisors, Robin Fencott, Jenny Attwater, Dan Stowell, Adam Stark, and my parents.

In addition to the above, I would like to offer thanks for advice, support, and discussing ideas to Simon Dixon, Geraint Wiggins, Mike Terrell, Andrew Robertson, Samer Abdallah, Matthew Davies, Chris Cannam, Sidney Fels, Andrew McPherson, Marcus Pearce, Andy Simpson, Jean-Baptiste Thiebaut and Charles Fox.

Thanks are due to Heidi Buehler, Jennifer Essex, Nicole Johnson and Tiff Chan for performing with the Serendiptichord, to Jan Lee and Tiff Chan for helping in the presentation of Impossible Alone and to others who supported these two projects in various ways: Di Mainstone of course, Dave Meckin, the Guerilla Science team, Jenny Attwater, Mark Ungar, Dierdre McCarthy, Rachel Lamb, Judy Zhang, Stacey Grant, Vesselin Iordanov, Richard Carvlin, Virginia Colman, Jenn Sheridan, Tom in San Francisco, Robin Fencott, Magdalena Chudy, Luis Figueira, Mike Terrell, Dan Tidhar, Boris Mailhé and Andrew Robertson.

I am also thankful of support from Becky Stewart, Mark Sandler, Gini Simpson, Joe Lyske, Mathieu Barthet, Andrew Nesbit, Melissa Yeo, Kevin Byron, Matthew Purver, Sharon Cording, Ho Huen, Daniele Barchiesi, Asterios Zacharakis, Martin Morrell, Chunyang Song, Saoirse Finn, Davy Smith, Henrik Ekeus, Thomas Wilmering, Matthias Mauch, Kurt Jacobson, Steve Welburn, Enrique Perez Gonzales, Katie Noland, Peter Foster, Stuart Battersby, other friends and my family.

Thank you to Ernest Edmonds, Sidney Fels, Cavan Fyans, Memo Akten, Ge Wang, and Massimo Boldrin for granting me permission to reproduce images and figures (noted in the text where appropriate).

This research was funded by a Doctoral Training Account from the Engineering and Physical Sciences Research Council. The Serendiptichord and Impossible Alone were additionally funded by the Platform Grant (EPSRC EP/E045235/1). The Serendiptichord was originally commissioned for the *ACM Creativity and Cognition Conference 2009*, produced by BigDog Interactive and supported by the Interactional Sound and Music Group.

Contents

List of Figures

List of Tables

List of abbreviations

Chapter 1

Introduction

> ...it is necessary to start to engage the audience in a way that can sustain interest for a noticeable period of time. The behaviour of the work needs to be interesting. The problem is identifying just what *is* interesting in interactive experiences.
>
> (Edmonds et al. 2006, p. 315)

1.1 Beyond marvellous

It is a familiar situation for many interactive artists. A work of magnificent complexity is created with a vision of an audience immersed exploring its endless possibilities. When exhibited, however, for those who participate it appears little more than a momentary marvel. In a museum, 5–10 seconds could be considered a long interaction (Edmonds et al. 2006). In contrast, both music and other interactive systems are able to captivate people for hours on end. For example, a videogame is often expected to last for 40–60 hours (Collins 2009). It does not seem to be the case that people are disappointed with a work and so move on, although undoubtedly this happens sometimes. The interaction simply appears to have been a fleeting distraction rather than an involving experience.

In this thesis, we will consider music-oriented interactive art—*interactive* *music*—and how it might be made more *captivating* over longer periods of time. There is evidence that interactivity has the potential to make sound installations more enjoyable (Gonzales et al. 2009). But whilst it is easy to gain attention, sustaining continued exploration remains elusive (Blaine 2006; Machover 2002). It seems that people lose their sense of wonder as soon as they 'figure out' a work (Rokeby 1998).

1.2 Interactive music

This thesis is about interactive music—a musical experience that involves participation from the listener but is itself a composed piece of music. In doing so the listener becomes a participant. Interactive music may take many forms such as a smartphone app or a sound installation exhibited in a public space. We will be focusing on interactive music experiences that arise when a participant interacts with a computer system, which we will describe as an *Interactive Music System* (IMS). Our enquiry will be limited to participants without any particular expertise who are interacting with a system for the first time. We will see a more specific definition of an IMS in Section 2.2.3.

1.3 Subjectivity

At the beginning of this chapter, a quote from Edmonds et al. (2006) describes the need to identify what it is about interactive experiences that makes them interesting, from which we may be able identify how a work might behave in an interesting way. We might have asked: 'interesting to whom?' The participant doing the interacting? The artist? An onlooker? However, the notion of an interesting interactive *experience* arrives with the assumption that it is the participant who we want to be interested.

The reason that we may ask this question is that 'interesting' is a subjective notion, by which we mean it is an attribute of a system that is measured based on an individual's perception rather than an objective measuring tool. We may sample a large number of individuals and aggregate their response in order to predict more reliably whether a further individual may find a system 'interesting'. From here we may extract general properties of systems that are likely to be found 'interesting'. These properties may be objectively defined and we may then be tempted to talk of them as if they were 'objectively interesting'. However, 'interesting' remains a subjective attribute. It does not make sense to talk of something being 'interesting' without an individual to find it so.

As we will see in Section 2.1, for many interactive artists, the final medium of the work is not the interactive system but the experience of interacting with it—a position we shall adopt throughout this thesis. Such an approach, where we are interested in the subjective experience of an interaction rather than any material outcomes, is referred to within the field of Human-Computer Interaction (HCI) as experience-centred design (Blythe et al. 2006). It has gained popularity within the study of computer-based musical interaction (Stowell and McLean 2011; Kiefer et al. 2008). However, often ideas that seem to begin with an experience-centred aesthetic vision result in abstract criteria of interactive artefacts without a detailed analysis of how these criteria will be perceived (e.g. Paine 2002).

Our interest in objective aspects of a system will be relevant only in terms of how they are manifest as subjective aspects through perception. Therefore, before stipulating requirements on how we want the system to behave we will need to consider what kind of experience to which they will be contributing. We will use the terms *interactive music* and *interactive music system* (IMS) to distinguish between this subjective experience we are trying to achieve and the system with which we interact to do so. The difference between between the two is analogous to the difference between audio and music.

Fundamental to designing an experience is understanding how people make sense of what they perceive (Wright and McCarthy 2010). Just as analysis of music without consideration of musical perception misses a great deal of what makes it music (Wiggins 2009), trying to understand an interactive experience without considering how an individual makes sense of it misses a great deal of what makes it an experience. Throughout this thesis, we will gradually draw in on what aspects of an interactive music experience might make it captivating and aesthetically rewarding, as well as how we might create it. Through modelling the process by which an interactive system is perceived, acted upon and understood, we may then consider how we may design IMSs to create such experiences.

For these reasons it does not make sense to begin with unnecessary objective limitations of interactive music system (IMS) is (cf. Rowe 1993; Paine 2002), or indeed what the difference is between 'reactive' and 'interactive (cf. Bongers 2006; Jordà 2005). We will, however, consider what we may learn from these approaches in Section 2.1.3.

1.4 Interdisciplinarity

This is an interdisciplinary work, spanning engineering, HCI, music, psychology, performance art and interactive art. It is important when doing interdisciplinary work not to become overly pledged to particular methods or ways of thinking. This is especially the case when mixing art and engineering as they are often at odds with each other when it comes to methodology, motivation, outcome and prejudices.

However, with that caveat, this PhD rests primarily as an engineering thesis for the following reasons.

- Our ultimate motivation is to identify how to produce specific effects (albeit subjective ones) rather than ideas or artefacts. Concepts, critical analysis and practical applications are important but remain tools to achieve this goal rather than ends in themselves.

- Theories that arise should be demonstrated through application and evaluated beyond personal introspection.

- The key contributions are theories, approaches and tools rather than artistic artefacts.

- These contributions should serve to assist others in producing similar effects.

- While it may be necessary to consider the immediate artistic value of how these contributions may be applied, it will be evaluated with respect to how it is received by its audience. We will not be looking for any further artistic or wider social significance.

  Note that this is not to say that these topics are not important or interesting; they are simply beyond the scope of this PhD.

Over the course of this thesis we will at different times have to be artists, scientists and musicians. But ultimately, as an engineering thesis, the goal is to *apply* these ideas to open up new possibilities.

1.5 Contributions

The primary contributions of this thesis are two complementary theoretical theoretical tools, the *Emerging Structures* model of exploration (Chapter 4) and a framework of *Perceived Agency* (Chapter 6), that assist in the understanding the subjective experience that arises through interacting with an IMS.

Along the way, we will make the following further theoretical contributions.

- We develop two design principles (Chapters 5 and 7) providing a general purpose means of applying the above theories in practice.

- We present a refined consideration of the composed instrument as a means of computer music performance distinct from ordinary Digital Musical Instruments (DMIs) (Chapter 3).

- The methodology available for evaluating the quality of a creative or aesthetic user experience is advanced. This includes a novel experimental design for investigating first-time use within a repeated measures design, created in collaboration with Prof Rosemary Bailey (Chapter 5), and refinement of Stowell's (2010) method of discourse analysis (Chapter 8).

- We apply the above theoretical tools to present a new means to classify types of interaction (Chapter 9).

In addition, the following practical works have been created during the course of this PhD.

- Impossible Alone, a full body interactive sound installation, created in collaboration with Tiff Chan (Chapter 7).

- The Serendiptichord, a wearable musical instrument for contemporary dance performance, created in collaboration with Di Mainstone (Chapter 3).

- The Manhattan Rhythm Machine, a screen-based interactive generative rhythm system (Chapter 5).

- Instigative Heads, an interactive sound installation created to accompany video work by the artist Alla Tkachuk, with support from Kurt Jacobson, Steve Welburn and Enrique Perez Gonzales (not documented in this thesis).

1.5.1 Publications

The following publications have arisen from this thesis (all based on material from Chapter 3).

T. Murray-Browne, D. Mainstone, N. Bryan-Kinns and M. D. Plumbley, "The Serendiptichord: Reflections on the Collaborative Design Process between Artist and Researcher," in *Leonardo*, **46**(1):86-87, 2013.

T. Murray-Browne, D. Mainstone, N. Bryan-Kinns and M. D. Plumbley, "The medium is the message: Composing instruments and performing mappings," in *Proceedings of the International Conference on New Instruments for Musical Expression (NIME-11)*, Oslo, Norway, 2011.

T. Murray-Browne, D. Mainstone, N. Bryan-Kinns and M. D. Plumbley, "The Serendiptichord: A wearable instrument for contemporary dance performance," in *Proceedings of the 128th Convention of the Audio Engineering Society*, London, 2010.

A. Otten, D. Shulze, M. Sorensen, D. Mainstone and T. Murray-Browne, "Demo hour," *Interactions*, **18**(5):8–9, 2011.

The following papers have been presented without corresponding publications.

T. Murray-Browne, "The Serendiptichord: Balancing predictable control with chance discovery in a wearable instrument for dancers," presented at the Sound, Sight, Space and Play postgraduate symposium, Leicester, 2010.

T. Murray-Browne, D. Mainstone, M. Plumbley, N. Bryan-Kinns, "The Serendiptichord: Balancing complexity with accessibility in a wearable musical instrument for dancers," presented at the Digital Music Research Network, London, 2010.

T. Murray-Browne, "How can interactive music engage audiences for longer?" presented at InterFace postgraduate symposium, London, 2011.

T. Murray-Browne, T. Chan, "Impossible Alone," presented at the Music HackSpace, London, 2011.

1.5.2 Performances and exhibitions

The following performances and exhibitions have arisen from the practical work involved in this thesis.

Instigative Heads Installed:

Mar – Apr 2009, Shunt, London.

The Serendiptichord Performed:

Oct 2009 ACM Creativity & Cognition Conference, Berkeley Art Museum, CA

Feb 2010 Kinetica Art Fair, London

Feb 2010 Swap Meet, The Barbican, London

Feb 2010 The Guthman New Musical Instrument Competition, Georgia Institute of Technology, GA

Jun 2010 INSPACE Gallery, Edinburgh

Sep 2010 Victoria & Albert Museum, London

Oct 2010 The Sweden National Touring Theatre

Feb 2011 Kinetica Art Fair, London

Impossible Alone Installed:

Jul 2011 The Secret Garden Party (with Guerilla Science), Huntingdon, Cambridgeshire

Sep 2011 Centre for Digital Music 10th anniversary event, Queen Mary, University of London

Nov 2011 Music Hackspace, London

1.6 The structure of this thesis

PhDs that mix art and engineering follow divergent pathways through theory, practice and evaluation (Edmonds and Candy 2010). In this thesis, the theoretical content has both guided and been guided by concurrent practical application and accompanying evaluation. Material is presented primarily in the order in which it was conducted. We iterate twice through the engineering style presentation of theory, application and then evaluation in Chapters 4–5 and Chapters 6–8. However, the actual development of this material was more interwoven than this. In particular, the outcome of the evaluation of Chapter 8 subsequently informed the theoretical content of Chapters 4 and 6.

This thesis bridges a large number of fields. As a result, there is a significant amount of background material to cover in Chapter 2. As well as providing context, there are a number of topics we will be building upon later in the thesis. We will indicate when this is the case.

Interactive music might be seen as a merging of the performance, creation and reception of music. There is a considerable wealth of research into all of these topics. Of particular relevance is the field of research into new musical controllers, referred to as New Interfaces for Musical Expression (NIME). As a result, we will consider quite closely the acts of performing and listening to music created with new musical interfaces through an exploratory project creating and performing a novel musical instrument, the Serendiptichord, in Chapter 3. The key finding from this chapter will be the importance of interaction design that *develops* throughout a performance, and the role that what the audience is expecting to see plays in their understanding and appreciation of a work.

In Chapter 4, we take the former of these lessons and consider how we may apply it within the context of interactive music. We derive a model of *Emerging Structures* to describe how an interactive experience may be continually evolving and developing without being confusing or inconsistent. Using this model, we unify a range of theories of exploration and musical perception. From this, we form a set of predictions as to what makes for a captivating interactive music experience.

In Chapter 5 we apply a number of these predictions to derive a general purpose design principle, *incremental constraint unlocking*, which is evaluated in a controlled user study. We find the principle does not have a statistically significant impact on user experience. However, a closer examination of the results as well as an informal analysis of some complementary qualitative data leads us to the conclusion that individual differences among the participants are instrumental to their response. In particular, some users seem to have preconceived ideas about how an IMS 'should' behave. We conclude that we need a more specific tool to be able to understand this attitude.

To develop this tool, we consider again the nature of non-interactive music in Chapter 6 to produce a framework of *Perceived Agency*. We justify this framework through applying it to a range of traditional and recently developed performance paradigms.

In Chapter 7, we use this framework to distinguish between *implicit* and *explicit* constraints within an IMS. From this we address a number of the problems with the above principle of incremental constraint unlocking with a new design principle, *the implicit constraint*. We implement this principle within a new IMS, Impossible Alone.

A rigorous qualitative evaluation of Impossible Alone is reported in Chapter 8. Note that although this is an evaluation of the ideas presented up to this point, the analysis it provided strongly informed our perspective on the earlier theoretical chapters.

In Chapter 9, we discuss the consequences of the preceding work and demonstrate the contribution it provides through addressing two questions: what makes for an interesting interaction and how can we describe it in a way that captures some of the richness of everyday experiences?

Finally, we conclude in Chapter 10 with a summary of the key findings, reflect on the research process of this thesis and outline a number of suggested avenues for further research that might continue this body of work.

We may think of the work presented in Chapters 1–3 as identifying what kind of experience we are trying to create with interaction music and forming the first part to this thesis. Chapters 4–5 and 6–8 are iterations of theory, application, evaluation forming parts two and three. Chapters 9–10 then form the final concluding part of the thesis.

Chapter 2

Background

Interactive music is a topic that lies at the intersection of a number of fields. The two that we will draw upon primarily are interactive art and New Interfaces for Musical Expression (NIME). these will be overviewed first (Sections 2.1 and 2.2). Following this, we briefly review some wider research on enjoyment and musical perception that we will be drawing on later in this thesis (Sections 2.3 and 2.4). Finally, we outline some of the evaluation methods in common use within the field (Section 2.5) before concluding in Section 2.6.

2.1 Interactive art

Interactive music systems (IMSs) fall into both the fields of interactive art and New Interfaces for Musical Expression (NIME). Both of these fields provide us with extensive research into how system design affects experience. In this section we review some relevant material from the former. NIME research will then be covered in Section 2.2.

Penny (1996) considers an *interactive artwork* as an interactive system that addresses artistic issues, where an interactive system describes 'a machine system which reacts in the moment, by virtue of automated reasoning based on data from its sensory apparatus.' Elsewhere, an interactive work is not considered 'complete' until it is interacted with. For example, Edmonds et al. (2006) and Rokeby (1998) both see the medium of the work as the experience of interaction rather than the machine by which this happens. This is analogous to the distinction we drew in Section 1.2 between interactive music and interactive music systems: the former is the experience arising through interacting with the latter. Therefore, whilst there are no hard and fast rules about what qualifies as *interactive art*, we may likewise distinguish between the interactive art experience and an interactive artwork as the artefact which allows this to happen. Typically, as with our IMSs (see Section 1.2), this will happen through the use of a computer and not require further assistance from its creator or another third party, and our scope of enquiry will be limited to these cases.

Whether or not it is interactive, any artwork will be perceived in a different manner by different individuals. However, interactive art extends this notion by invoking a spectator to act based on their perception of the work, which then leads to an altered perception. The spectator must take a more active role in order to perceive the work, exploring its behaviour through experimentation. Correspondingly, the artist must find a way to express themselves creatively whilst leaving the ultimate realisation of their work open (Tanaka 2007). If others in the audience are present, then through interaction the participant's response to the work is visible and the process of viewing art becomes a more social and shared experience.

Rokeby (1998) describes the creation of interactive art as the *construction of experience*, experience that involves making decisions and doing something rather than experience in which content is simply fed to the audience. Irrespective of debates over the definition of interactive art, many interactive artists argue that the value of the work is within the interaction itself rather than the response of participant or system considered in isolation (Edmonds et al. 2006).

Examples of interactive art range from public space installations to websites and smartphone applications. We may question where we draw the boundary. For example, when are smartphone apps considered 'interactive artworks'? We avoid prescriptive definitions by simply relying on how the creator describes their work, which we assume indicative of their intent.

Our area of interest will typically be work where an audience may be inquisitive but has not been trained and is not significantly motivated to use the system by factors beyond the potential experience they might have with it. This context is common to work installed within a public space (Paine 2002). A crucial consideration of such works is how their audience will learn to use the interface, given the limited amount of time participants will invest in doing so and how diverse they are likely to be in terms of behaviour and familiarity with interactive systems (Blaine 2006; Dipper 2009)—although not all interactive artists necessarily seek to create a work that is 'learnable' (Boden 2005).

2.1.1 The perceptual feedback loop

Fundamental to any conscious interaction between human and machine is the perception/action cycle, or *perceptual feedback loop*, which describes the process whereby the output of an interface provides the user with details of the result of their input (Sales Dias et al. 2009). This allows them to evaluate what they have done and plan further actions and in doing so establish a feedback loop (Norman 2002) (Figure 2.1). A simple example of this is a console application, where characters entered into the interface are displayed onscreen.

Through this process, a user is able to form a *mental model* of how the machine works. Our mental model is a cognitive representation that allows us to reason about the consequences of our actions. It need not be a true representation of how a system actually works, just accurate enough to allow us to form plans and achieve our goals (Norman 2002). Such models may be formed both consciously and unconsciously (Preece et al. 2011, p. 86). The exploratory process by which they are established will be considered in detail in Chapter 4.

The perceptual feedback loop describes the process through which an

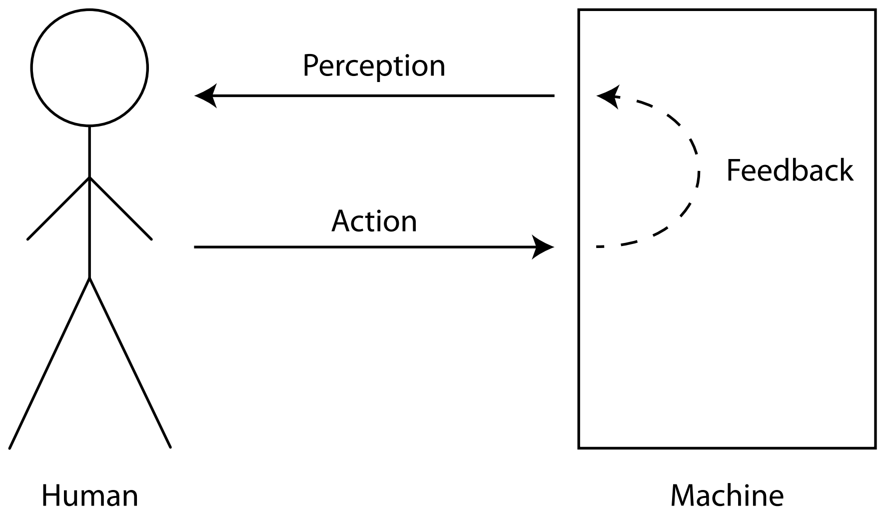
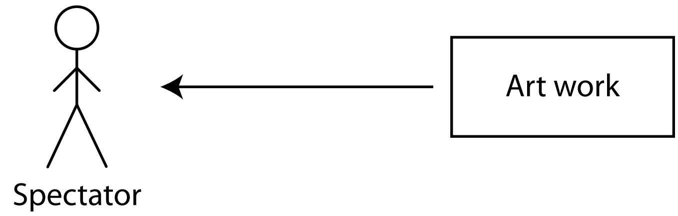


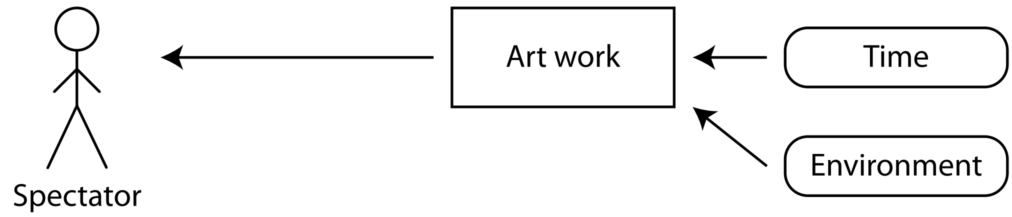
Figure 2.1: The perception/action cycle, or *perceptual feedback loop*. The user is able to construct a mental model of how the system works through perceiving the feedback that results from their actions.

individual becomes aware of how they may exert control over their environment. In Section 2.2.11, we will review literature that considers this issue in terms of how reliable, diverse and accurate a sense of control this feedback loop provides. Later on, in Section 6.3, we will introduce the more specific concept of agency to describe the exertion of intentional control over a process.

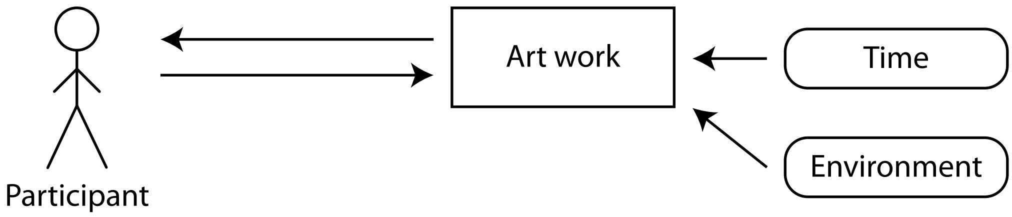
In the remainder of this section, we will outline a number of theories that attempt to understand in more detail the experience of interactive art. As well as drawing upon these ideas later in the thesis, they provide the context within which our work is created.



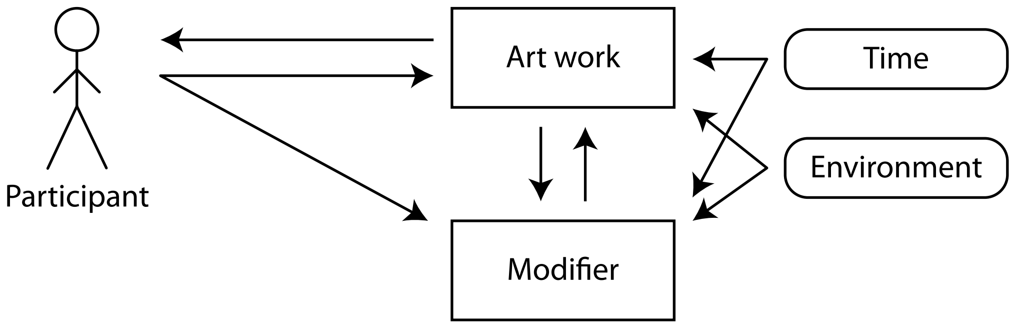
(a) The static system



(b) The dynamic-passive system



(c) The dynamic-interactive system



(d) The dynamic-interactive system (varying)

Figure 2.2: Cornock and Edmonds's classification of process-oriented art. Adapted from Cornock and Edmonds (1973).

2.1.2 Cornock and Edmonds's process-oriented art

Cornock and Edmonds's (1973) provided one of the earliest classifications of interactive art within their framework of *process-oriented art*. Taking a systems-theory approach, they consider how influence flows between the artwork, an individual member of its audience, a 'modifying agent,' time and the environment. They provide four categories (Figure 2.2).

1. **The static system** refers to any unchanging work of art and includes most traditional works of art. It is simply spectated by the individual.

2. **The dynamic-passive system** changes in response to time or the environment but not in response to the spectator.

3. **The dynamic-interactive system** changes in response to the individual, who is now referred to as a participant. A feedback loop is established.

4. **The dynamic-interactive system (varying)** is a special case of system **3** where 'an artist modifies the system or process in a way not allowed for in its original definition.'

Different members of the audience may be involved in different systems that include a single artwork. For example, if one member of the audience has influence over the work and another does not then the former will be in a dynamic-interactive system whereas the latter will be in a dynamic-passive system as the participant is considered an environmental influence from the spectator's perspective.

There are a number of problems with this model.

The definition of system **4** is unclear. What is the 'original definition' of a system? Edmonds et al. (2006) expand the definition to allow the modifier to be a human or software process stating that the resulting interaction will depend on the history of interactions with the work and is thus not predictable. But they still refer to the 'original specification' of the work. From this should we infer that system **3** is predictable and thus deterministic and non-chaotic? Can the artwork be its own modifying agent? Clearly not, otherwise that would have been allowed for in the original specification. Should we infer that the output of system **3** does not depend upon earlier input from the participant? Would this then disallow simple signal processing such as smoothing a gestural input?

The classification is oriented around the artist's perspective. An audience member may be within system **4** from the artist's perspective but unaware of the modifying agent and so think themselves in system **3**. If the individual does not realise that they are influencing the artwork then there will be a feedback loop but no *perceptual* feedback loop (Section 2.1.1) and they may think themselves in system **2**.

There is an unstated presumption that 'normal interaction' is a learnable well-defined specification relating input/output pairs whilst 'varying interaction' ranges in scope from the artist correcting a bug in the software to an interaction design that evolves with such speed and complexity that it appears completely random. Here the artist would observe system **4** but the audience system **2**.

In spite of these shortcomings, the motivation behind the model seems similar to ours. Cornock and Edmonds (1973, p. 13) write of system **3**: 'this system can be very rich, though the speed with which the participant may exhaust the set of possibilities means that the result could lack substantial interest or value.' The implication being that system **4** by varying will continually have something original to offer keeping the audience interested. In Chapter 4, we will examine this idea in more detail.

There is also a strong insight underlying the model. It aims to classify not through attributes of the artwork itself, but the relationship between work and audience. Someone may move through different types of system as they interact with a single work and there may be multiple systems of interaction happening simultaneously with different members of the audience.

2.1.3 Definitions of *interaction*

What should be meant by the term *interaction* has received considerable debate within the field. Throughout this thesis, we will use the term to describe any process in which user and system are influencing each other's behaviour in a perceptual feedback loop as *interaction*. A system capable of being a part of an interaction is hence described as *interactive*.

Others have presented more stringent requirements on what interaction 'should' mean. It is not our intention to debate the meaning of words—our focus is experience rather than classification. However, we will review some definitions proposed by others as it will inform our understanding of what types of experience others are trying to create with their systems.

Dannenberg and Bates (1995) presents a model of interactive art with loosely defined requirements as an agent of the artist. He 'looks for' systems that

- *perform a substantial amount of decision making and generation of artistic content*,
- *can perform rudimentary perception, create internal models of their users, plan future activities, and call upon stored concepts*,
- *have large amounts of memory giving rise to complex behaviour*,
- have a *significant interaction component* (Dannenberg and Bates 1995, p. 5).

Dannenberg and Bates acknowledge that these requirements are on a spectrum leaving no clearly defined boundary of what counts as interactive art. However, note that there is a blend between unquantifiable subjective requirements (amounts of decision making, the significance of the interaction component, the complexity of the behaviour, the extent to which content is artistic) and objective requirements that could well be imperceivable to a participant (perception, user models, future plans, stored concepts).

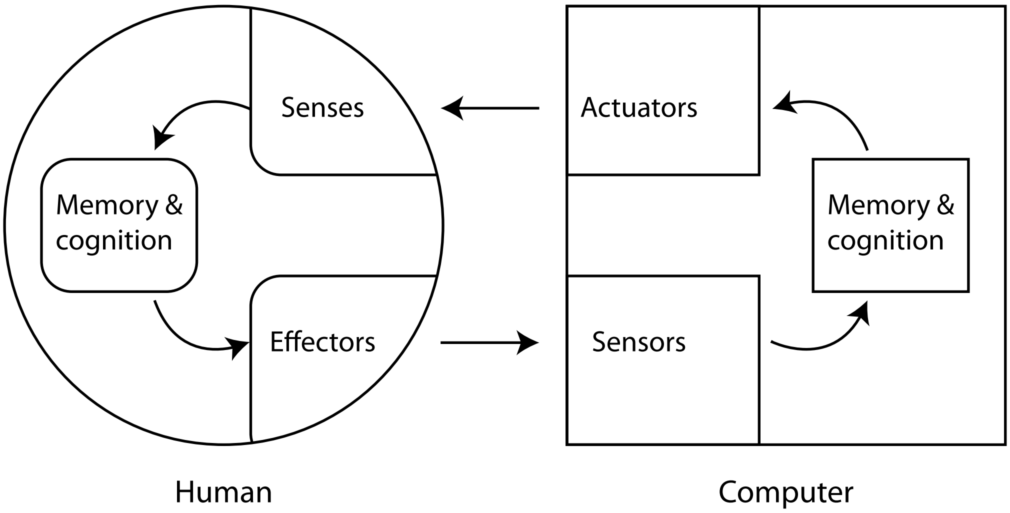


Figure 2.3: Bongers's human-computer interaction loop with the requirement of memory and cognition. Adapted from Bongers (2000).

Bongers (2000) describes human-machine interaction as a feedback loop between human and machine requiring sensing, memory, cognition and actuators from both parties (Figure 2.3). Without cognition, for example, it is merely *reactive*. Bongers believes that 'ideal' interaction should be *mutually influential*, which according to Bongers (2006) means that following the interaction both human and machine are left in a different state, frame of mind or with different views. But he does not justify why this might be ideal, why we would like our machine to think and remember or why the state of the machine following the interaction should matter.

Jordà (2005) similarly defines interaction as a process of mutual influence where information from each party influences the behaviour of the other. 'True' interaction requires extensive two-way influence. However, common use of real-time signal processing effects, for example, he deems lacking in Human→Machine interaction as the effects that are applied are independent of the performer. This is a strange objection: even without influencing which effect is applied, the entire output of such a system is a transformation of the input and therefore quite strongly influenced by the human. The effect may also require, as Bongers desired, significant cognition and memory. There appears to be some unspoken assumptions about exactly what needs to be influenced.

Interactive artist Paine (2002), like Bongers, argues that without a level of cognition a system (and we have moved now from considering a process to considering the system in isolation) is reactive. However, Paine then provides the example of a racing arcade game, which he deems non-interactive because the *racetrack* is not changing in response to the user. His requirement is then refined to demand that the entire history of inputs be influential on each output and that the system promises continually new outcomes.

In each of these cases, although mutual influence is stated as the requirement, there appears to be a hidden demand that this influence is not with the system but some entity *within* the system. This is an echo of Cornock and Edmonds (1973) *modifying agent* that alters a system's 'original specification' (Section 2.1.2).

These discussions assume that the difference between interaction and reaction is an objective reality determined by properties of the machine. But from the participant's point of view (i.e. subjectively), the only means by which any change in the system may be determined is through *how* input affects its output. The modifying agent with whom we are interacting is completely imperceptible except through its influence over this relationship.

Chadabe (2005) provides three metaphors to describe interaction (Drummond 2009).

1. The powerful gesture expander (a deterministic system that translates simple inputs into complex musical outputs).

2. Sailing a boat on a windy day and through stormy seas, where precise control is not assured.

3. The conversational model, where both party contributes but neither is in control of the final outcome.

Bongers, Jordà and Paine all refer to *conversation* or *dialogue* in their attempts to describe what makes some feedback loops more interactive than others, Paine most explicitly. Therefore, we may assume that they all fall within Chadabe's conversational metaphor.

Johnston et al. (2009) categorise interaction as being *instrumental*, *ornamental* or *conversational* depending on whether control of the sound output is respectively with the participant, the system or shared between the two. Conversational interaction is described as rapidly shifting between instrumental and ornamental interaction.

The term *conversational interaction* implies that there is something of the conversational experience that we would like to have when interacting with our system. As the distinction between Chadabe's latter two metaphors imply, there is something more that mutual interaction. However, as evocative as Chadabe's metaphors are, it is unclear how they might be practically realised. Even from an abstract perspective, there are numerous possible distinctions that we might draw between sailing from conversing. Which ones are important? After we have defined the two fundamental theoretical frameworks of this thesis—*Emerging Structures* in Chapter 4 and *Perceived Agency* in Chapter 6—we will be able to provide a more concrete answer to this question in Section 9.4. In particular, we will justify the point of view that conversational interaction is necessarily a subjective experience that cannot be defined without a consideration of the participant's *perception* of the interaction.

For these reasons we have avoided including any loosely defined subjective requirements such as complexity or memory within the above definition of *interactive* and use it simply in the objective sense outlined above to describe a system that establishes a feedback loop with a participant. How interaction is subjectively perceived is a topic that we will build up in detail throughout this thesis.

Figure 2.4: A participant standing in front of *Absolute_4.5* (2006) by Ernest Edmonds and Mark Fell. Photo: Greg Turner. Reproduced with permission from Candy and Edmonds (2011).

2.1.4 Edmonds et al.'s model of engagement

Edmonds et al.'s (2006) model of engagement draws on observational research of how visitors engage with interactive exhibits in museums. It is explored in relation to Edmonds and Fell's work *Absolute_4.5* (2006) where solid bars altering in colour and position are projected onto a screen with synthetic sounds driven by the same algorithmic process (Figure 2.4). Participants influence the behaviour of the work through hidden pressure sensitive floor pads providing a two dimensional input.

The framework considers how a work establishes a relationship with its audience through three stages: *attracting* interest, *sustaining* interest and *retaining* interest beyond the interaction.

Attractors are the aspect of a work that attract potential participants to it in the first place. These will be aspects that catch the attention of passers by. In *Absolute_4.5*, this was initially dealt with by setting the resting state of the work, where nobody was interacting with it, to be noisy and flashing (a tactic similar to those found in videogame arcades (Collins 2008)). This was effective, but detracted from other works in the space and so had to be adjusted to find a balance.

Sustainers are aspects of the work that keep an audience interested once it has their attention. In museums, exhibits that have a strong holding power lead to what Bollo and Dal Pozzolo (2005) describe as *hot spots*. The behaviour of the work has to be *interesting* and it is identifying what exactly *is* interesting in interactive experiences where the difficulty lies (Edmonds et al. 2006). With *Absolute_4.5*, although loud and active in its resting state, the work quickly became subdued and quiet once a participant entered its space. This was an *initial* sustainer through puzzling the participant about what would happen next. Within the space, the work then responded to two dimensions of input, one reacting in an immediately perceptible manner and one more subtle, giving the audience a combination of something they could quickly grasp with something more difficult to work out.

Relaters are aspects of a work that establish a continuing relationship with a participant beyond the interaction, potentially leading to them returning or recommending it to others. *Absolute_5*, a developed version of the above work, investigated relaters somewhat through implementing a long-term evolution of the means by which it responded to its participants, as well as also responding to participants in a different city. Edmonds et al. (2006, p. 319) write that the concept behind this work was to create an interactive experience that was unique each time it was visited, and that 'The changes are not random, although perhaps they might be thought to be, but are influenced deterministically by people out of sight in another city.'

The topic of our investigation is primarily in the middle category of the work—the *sustainers*. The framework is useful in highlighting other aspects of a work that should be considered in many contexts. Attractors and relaters are typically overlooked in lab-based research although for the latter we might reasonably assume that a long captivating experience will form a stronger relationship with a participant than a brief momentary marvel. However, the museum-based background of the model carries assumptions of works competing for attention much like in a videogame arcade, which may not necessarily be the case.

Edmonds et al. description of the sustainers in *Absolute_4.5* is a pattern we shall meet in a variety of forms throughout this PhD—a puzzle provided to its participants in combination with an indicator that there is a solution to be found.

2.1.5 Rokeby's construction of experience

Rokeby (1998) sees interactive art as an extension from simply providing content to a spectator as also specifying the means by which that content may be experienced. The interface to the work privileges certain paths of exploration, sometimes explicitly but often more subtly by making some actions easier and more obvious. This leads to what Rokeby dubs *operational clichés*—mechanisms that are so neat and easy that they become overused.

Any interface embodies a model of the user in terms similar to the way we model systems: what they may perceive (their input) and what inputs they may provide to the system (their output). In order to use the system, the user must internalise this schema to some extent and this modifies their sense of self within the installation. Rokeby provides an example how this distorted sense of self may manifest with his work *Very Nervous System* (Rokeby 1986).

The *Very Nervous System* is a sound installation that responds to movement. The system receives input from an 8x8 pixel camera and calculates a Fourier transform on each pixel, which trigger different sounds. This effectively projects a fixed grid onto the space, with a participant able to control the sound through varying the location, amount and speed of their movement (Rokeby 2011; Winkler 1997). Rokeby (2011) found that participants would begin to respond unconsciously to the system with reflexive movements and this would establish a feedback loop. However, the participant's consciousness of their movement would 'lag' behind these reflexes leading to a perceived loss of conscious control over one's physical actions.[1] Rokeby (1998) describes the effect as intoxicating, addictive and leading to states

[1]This sensation is referred to as losing a *sense of agency* within psychological literature (Moore et al. 2009). However, we will be using the term *agency* extensively in Chapter 6 to mean something slightly different so we will avoid this specific use of the word.

that could be described as 'shamanistic'. At the other end of the spectrum, reducing the dimensions of interaction available to the participant can make their impact on the system more immediately recognisable. This, paradoxically, can increase their sense of empowerment (Rokeby 1998).

Rokeby's reflective research practice provides some compelling insights into his diverse artistic practice. He highlights the way an interface can lead users along certain pathways without establishing explicit limitations as well as the crucial role of the perceptual feedback loop in establishing a sense of control over events. However, as an artist he provides examples what can be achieved and what the consequences are, rather than general principles of how to go about achieving it.

2.1.6 Fels's model of embodiment

Embodiment describes an approach within HCI that considers the role of the body in perception (Larssen et al. 2006). It is grounded in the notion that our understanding of abstract interaction is derived from our familiar understanding of interacting with the physical world (Preece et al. 2011, p. 96).

The term is also used to describe a relationship between user and tool in which the tool drops from the user's conscious awareness and they instead perceive themselves as directly interacting with the final medium (Fels 2000). For example, computer users quite quickly stop talking about 'pressing the mouse button' and instead refer to 'clicking on the screen'.

Fels (2000, 2004) argues that embodiment can also be perceived the other way, where the user sees themselves as being manipulated by a system. This leads to a model of how a person relates to an object of four categories as follows.

Response. The person is disembodied from the object and communicates in a dialogue. Any pleasure from the relationship derives from the result of their interaction rather than the interaction itself.

Control. The person embodies the object and feels like it is an extension of their own body and mind. In this situation, pleasure arises from the relationship itself rather than its outcome. This relationship is similar to that between an experienced musician and their instrument.

Contemplation. The object communicates to the person. This situation is similar to traditional forms of media where there is no interaction. The person may respond to the signals coming from the object through introspection.

Belonging. The object embodies the person. In this category, the person derives pleasure from a relationship where they have relinquished control to the object. Fels (2000) sees this final category as rich for experimentation.

This model arose through Fels observing participants progressing through these different stages (excluding the third) with his interactive artwork *Iamascope* (Fels and Mase 1999). *Iamascope* is an installation consisting of a video projection, sound system and camera positioned to capture a participant looking at the projection. The projected video is created through cropping and reflecting the camera feed into sectors of a circle in a manner similar to a kaleidoscope (Figure 2.5). To create the sound, the same cropped camera feed is divided into zones. Each area is associated with a particular pitch selected from an automatically cycling list of chords. Notes of a fixed duration are triggered when the average intensity over a zone exceeds a threshold. Whilst musical, it is a system created for an aesthetic experience rather than expressive musical performance (Fels et al. 2002).

Fels identifies a number of aspects of *Iamascope* that lead participants to progress into this fourth category of embodiment, such as *mirroring* their body in the transformed image whilst not being abstracted so far that it becomes unidentifiable. Elsewhere, he has identified metaphor and collaboration as additional tools that may lead to this type of experience (Fels 2004). This notion of the participant relinquishing control is similar to the

Figure 2.5: A participant interacting with *Iamascope* (1997). Reproduced with permission from Fels and Mase (1999).

loss of agency reported by Rokeby (1998) in Section 2.1.5, as is the proposed cause: a feedback loop that is very sensitive to the participant's body.

Costello et al. (2005) conducted three in depth qualitative studies of *Iamascope* combining video-cued recall[2] with video-based observation which was then coded by a multidisciplinary team (Bilda et al. 2006). They found evidence supporting Fels's categories of embodiment as well as more detailed subcategories.

1. **Response** was observed to progress through two subcategories:

   How to work it, where the participant tests the broad parameters of the system with few expectations about the outcome,

   Realisation, when the participant recognises their face within the image making the underlying concept behind the mapping clear.[3]

2. **Control** through three subcategories:

   Purposeful movement where the participant moves with more confidence and commitment and has expectations about the outcome,

   Fine and detailed movements where the participant investigates more nuanced aspects of the mapping and gains satisfaction from exercising creative and expressive control,

   Goals and aims where the participant not only has expectations regarding the outcome but goals and aims that they are trying to carry out.

3. **Contemplation** was accompanied by the participant relating the work to other parts of their life such as who they might discuss their experience with.

[2]Video-cued recall is a research method where participants are invited to be videoed interacting with the system and then provide a running commentary whilst watching the video. See Section 2.5.5.

[3]*Mapping* is defined in more detail in Section 2.2.7.

4. **Belonging** was identified in one participant, who was strongly engaged with the installation for longer than the others.

Costello et al. observed that Realisation typically marked the start of a transition from Response to Control and describe a shift in attitude from *investigative exploration* ('what does this object do?') to *diverse exploration* ('what can I do with this object?'). They observed that participants would shift between these states, returning to the response state when they were bored.

They also proposed a fifth stage Disengagement, observing that before decided to stop using the system each participant had repeated an action sequence they had recently done. This corresponded with participants descriptions of *exhausting the possibilities*.

Fels's model provides a useful model of the process of engagement a participant might follow with an interactive system. In particular, we will review in more detail Costello et al.'s identification of alternations between investigative and diverse exploration and the issue of exhausting the possibilities in Section 4.1. However, even with the fifth category of disengagement, it remains an ideal model focusing on what can go right. It does not account for frustration with how the installation responds or goals that could not be achieved. It provides some indication of how to achieve the different states, but it is not entirely clear how easily these would generalise to other works. Furthermore, although Fels (2004) acknowledges that the states are not mutually exclusive, Contemplation does not seem to fit into the model with the other states which themselves do seem somewhat more mutually exclusive.

2.1.7 Reeves et al.'s model of the spectator experience

Reeves et al. (2005) created a two-dimensional taxonomy of spectating interaction by considering the extent to which the *manipulations* and the *effects* of the interaction are visible. Each axis ranges through hidden, partially revealed or transformed, revealed and amplified.

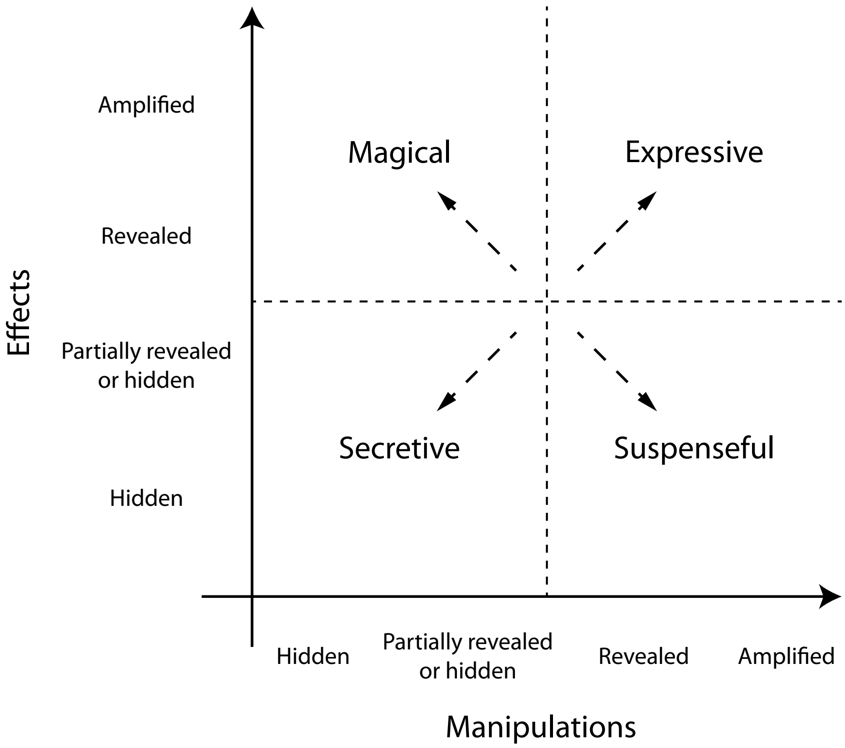


Figure 2.6: Reeves et al.'s model of the spectator experience. Four quadrants describe the quality of an interaction depending upon the extent to which the manipulations and the effects are revealed. Adapted from Reeves et al. (2005).

Reeves et al. establish four categories of interaction from the four quadrants of the graph (Figure 2.6).

Expressive interactions have strongly visible manipulations and effects. These allow for the understanding by spectators of how a system is being interacted with. In performance this may be to appreciate skill. With an interactive installation it may serve the role of allowing spectators to form an understanding of how to interact with the system before they use it.

Secretive interactions have mainly hidden manipulations and effects. With an interactive installation this may prevent surprises within the system from being revealed to future participants before it is their turn.

Magical interactions have hidden manipulations from a visible performer but clearly visible effects.

Suspenseful interactions have visible manipulations but no visible effects. Spectators may be attracted by seeing the interaction and drawn into participating in order to see the effects.

Sheridan, Bryan-Kinns and Bayliss (2007) argue that the dichotomy between performer and spectator within this model does not capture the intricacies of how spectators *become* performers—we will see how they address this issue below. However, for us there is a more fundamental issue with this model. Although the model describes *what* an audience may perceive, it does not consider *how* they perceive it. For example, as we will argue in Section 6.1.4, magical interaction may well be interpreted as *automation*. Suspenseful interaction may simply be perceived as an error. The perception of error will be discussed in more detail in Section 2.2.14

2.1.8 Sheridan's framework of witting participation

Sheridan (2006) presents a framework of Witting Participation to describe the multitude of roles an individual may take within a performance. Her roles are described in terms of the *performance frame* which defines the cognitive context defining the rules, interpretation and behaviour of an activity. Behaviour is characterised along three dimensions.

Wittingness: The extent to which an individual is aware of the performance frame.

Technical skill: Whether the individual demonstrates skills that are relevant to the performance frame.

Interpretive skill: The extent to which an individual may use their technical skills to convey meaning.

Sheridan (2006) argues that an individual progresses from being a *bystander* into a *spectator* when they become *witting* and choose to engage with the performance frame. The extent to which they may then transition to the role of participant is likely to depend on the extent they are able to acquire the necessary technical skills as a spectator. However, a participant remains separate from a performer, who possesses interpretive skills and exercises control over the performance frame itself.

The framework is illustrated through two related interactive works *iPoi*, (Bayliss et al. 2005; Sheridan et al. 2006a,b, 2007), later developed as *uPoi* (Sheridan and Bryan-Kinns 2008; Bryan-Kinns et al. 2010).

Poi is a type of performance art where a weight attached to a length of cord is held in each hand and swung around the body. The term *poi* also refers to the objects that are swung. *iPoi* and *uPoi* are poi augmented with embedded accelerometers that communicate wirelessly with a nearby laptop, which produces an audio and visual response. Originally created for poi performers within nightclubs with the intention of extending the impact of the performance without detracting from the performer's ownership or focus (Bayliss et al. 2005), later developments focused on assisting transitions from spectator to participant (Bryan-Kinns et al. 2010). They included, for example, extra sensors not embedded within the poi to allow non-performative participation (Sheridan and Bryan-Kinns 2008).

iPoi and *uPoi* differ from *Iamascope* or the *Very Nervous System* in that they are created with a focus on participants who already have a specific performative skill. Later developments allowed for multiple sets of *iPoi* to run simultaneously as well as a sensor placed inside a teddy bear allowing less performative interaction for shy participants (Sheridan, Bryan-Kinns, and Bayliss 2007).

We will explore in more detail performative ownership and the role performance context plays in perceiving skill when we develop a framework of Perceived Agency in Chapter 6.

2.1.9 Jo and Tanaka's matrix of musical participation

Jo and Tanaka (2009) present a framework with three levels of participation considered over four areas of a musical activity. Participation is categorised in terms of an individual's ability to influence the outcome of an area and based on Arnstein's (1969) model of citizen participation. They are as follows.

Non-participation. The individual is unable to influence the outcome

Tokenism. The individual is somewhat able to influence the outcome but power to change the status quo rests with others.

Citizen power. The individual is able to negotiate a change of outcome with others.

The four *perspectives of sound making practice* over which they consider an individual's permitted level of participation are

Sound: Any attribute of the sound produced from their instrument including pitch, temporal shaping, volume, timbre.

Instrument: The set of objects that people manipulate to create sounds.

Process: The sequence of actions taken by those present during the participation.

Performance: The potential for linking listening to others with playing and so forming a social exchange.

Jo and Tanaka consider a number of different musical performance contexts in terms of these categories including the traditional orchestra concert, a drum circle and a number of experimental musical works that stipulate experimental performance practice.

Somewhat inspired by a desire to explore new forms of musical practice, the matrix encompasses a broad spectrum of musical activities. It is useful to our needs through highlighting the importance of considering interactivity—analogous to level of participation—beyond the audio output of a system.

However, for the context of our enquiry, it is perhaps somewhat coarse. For example, as every aspect of sound falls into a single category, it cannot differentiate an improvised melodies from an expressive performance of a fixed score. It also does not consider how one's participation levels might change throughout an activity as does, say, Sheridan's (2006) model (Section 2.1.8). In Chapter 6 we will return to these topics with a greater focus on authorship, intention and performance context.

We have covered a few selected theories regarding interactive and performance art that consider what is perceived, what kind of experience it leads to and the role both an interactive system and a performance context may play in determining the participatory role taken by those present.

Most of the work has identified what kind of experiences an audience member may have and how aspects of a system and the context in which it appears may influence this. In Chapter 4, we will attempt to consider much more finely *why* the interactive experience itself unfolds as it does by drawing on the psychology of musical perception, exploration and motivation. In Chapter 6, although we derive it independently, our framework of Perceived Agency will in particular build upon the frameworks outlined by Sheridan et al. (Section 2.1.8) and Jo and Tanaka (Section 2.1.9).

2.2 New Interfaces for Musical Expression

The research field and community of those exploring new means to transform human action into sound is often referred to as *New Interfaces for Musical Expression* (*NIME*)—named after the key conference in this area.[4]

As mentioned in Section 1.2, we are using a specific interpretation of the term Interactive Music System (IMS). The meaning of the term *instrument*

[4]The International Conference on New Interfaces for Musical Expression (`http://www.nime.org`. Accessed 27 March 2012).

is also often debated (e.g. Jordà 2004). We will therefore use the term *Digital Music System* (DMS) to cover the entire range of systems discussed in this chapter.

2.2.1 A brief history of new musical interfaces

The Theremin (1928) is often given as a starting point in the history of experimental electronic musical instruments (e.g. Roads et al. 1996, p. 622). Whilst not the first instrument to be powered electrically rather than through human effort (organs were being electrified as early as the 1900s (Sefl 2006)), it diverged significantly from previous instruments. Firstly, its sound was created from oscillators rather than the vibration of physical materials interacting. However, arguably more significant within the field of NIME was the interface through which its performer controlled the instrument. With two sensing plates capacitively coupled with the performer's hands, pitch and amplitude were each controlled continuously and independently by one hand without any contact with the instrument (Paradiso 1997). Throughout the mid-20th century, further electronic instruments were developed, often using interfaces derived from the piano keyboard such as the ondes Martenot and the Mellotron (see Chadabe (1997) for a detailed overview).

From the 1950s onwards, modular synthesisers became commercially available—most notably from companies founded by Robert Moog and Don Buchla—consisting of separate signal generating units that were connected together through plugging in cables known as patch cords. Although early modular synthesisers typically had keyboards attached, the synthesiser itself required time and expertise to patch—neither of which needed to be the performer's. This opened the door for new types of relationships between performers and instruments: an instrument that one knows how to play but not program, or vice versa. The possibilities for programmable instruments magnified in the 1980s with the introduction of the Musical Instrument Digital Interface (MIDI) protocol, an industry standardised communication protocol that allowed an arbitrary controller (often either a keyboard or a sequencer) to be easily and quickly connected to an arbitrary synthesiser (Miranda and Wanderley 2006, p. 19).

Within the past two decades, however, we are within something of a renaissance within the field. A huge range of novel musical interfaces are currently being created by professionals and amateurs around the world. This has been driven largely by the availability of powerful computers bringing with them cheap software synthesisers as well as affordable sensing technology. Just about any kind of existing sensor that measures some aspect of a performer's gestures can be (and quite likely has been) adapted as a musical interface. Examples include the microphone (e.g. Stowell and Plumbley 2010), the webcam (e.g. Lyons et al. 2003; Kiefer et al. 2009) and accelerometers (e.g. Magnusson 2010). Most of these systems perform a substantial amount of processing to reduce the huge amount of sensor data into something more manageable.

There are two key paradigms within NIME research that are most relevant to our research: the Digital Musical Instrument (DMI) and the interactive sound installation/application. We discuss both of these below, before considering some other paradigms of interest in Sections 2.2.4 and 2.2.5.

2.2.2 Digital Musical Instruments (DMIs)

In recent years, the *Digital Musical Instrument* (DMI) has been the most dominant paradigm within NIME research. A DMI is defined by (Wanderley 2001a) as an instrument that divides into a separate gestural interface and sound generation unit with a *mapping* to specify how the interface controls the generator. Miranda and Wanderley (2006, p. 1) define DMI similarly with the further stipulation that a computer is used to generate the sound. DMIs are created as musical *instruments* and thus arrive with the assumption that they are to be used in the same context as a conventional instrument might

(Gurevich and Fyans 2011). However, in spite of the extensive amount of research into DMIs, there is a sense of dissatisfaction that those performing live computer music still rely on 'generic and dull' MIDI controllers (Jordà 2004), the laptop keyboard (Widmer et al. 2007) or are simply unsatisfying to watch (Schloss 2003).

Jordà (2005) distinguishes a range of characteristics that any instrument should have, many of which we will cover in more detail later in this section. However, he argues that instruments differ from other interactive music systems in that they permit

- improvisation,
- experimentation,
- the performance of a range of different pieces of music.

The division between input and output has been referred to as *splitting the chain* (Jordà 2004). It frees those creating instruments from physical constraints but at the same time can lead to losing fundamental physical traits such as tactile feedback (Wanderley 2001a). There is no longer the *feel* of creating sound (Widmer et al. 2007). Furthermore, a problem of *coherence* arises with controllers or synthesis engines created independently and then 'glued' together at the last moment. The physical origins of controlling conventional instruments are intuitively felt by performers and spectators alike, even if they are not explicitly understood. However, a general purpose 'musical controller' will not necessarily relate to the sound it controls in a meaningful way (Jordà 2004). Both Fels (2004) and Jordà (2005) have cited the fact that few DMIs persuade musicians beyond their original creators to invest time to learn and master them as a sign that they are not being made to satisfaction.[5]

[5]This state of affairs may be changing. There is evidence instruments such as the Reactable, the Ocarina and the Eigenharp (all discussed below) have begun developing their own dedicated performers.

Figure 2.7: The Reactable with a number of its marked objects above the translucent surface onto which the visual feedback is projected. (Photo: Massimo Boldrin. Reproduced with permission.)

Examples

Space limits us from providing extensive details of existing DMIs. However, we will briefly describe a few examples to illustrate some of the diversity within the field.

One of the most successful instruments in recent years is the Reactable (Jordà et al. 2005). The instrument consists of a round translucent table. Underneath the surface, a projector, infrared camera and computer turn this into a large touch screen. As well as touches, specially marked objects placed on the table can be tracked with respect to their position and orientation. Each object acts as a module as part of a modular synthesiser transmitting either control or audio data, with proximity between objects determining whether or not they are connected. Crucial to the Reactable is the projection, providing instantaneous and rich feedback about how the physical location of the objects is being interpreted by the system (Figure 2.7). Although it is also used in performance as a solo instrument, the Reactable was designed from the outset to be a *collaborative instrument* allowing a group of musicians to create music within a shared space (Kaltenbrunner et al. 2006).

Some DMIs are created in imitation of traditional instruments, though usually extending their capabilities somewhat. This can allow performers to draw on established technique (Cook 2001, 2009). It may also assist audiences in understanding the interaction that they are spectating.

Don Buchla's Marimba Lumina (Buchla 2005) is a controller with bars laid out in a similar manner to the marimba. It is played with colour coded mallets containing tuned coil, each of which are detected electronically by the bars, allowing the instrument to detect velocity, position along the bar and which mallet hit it.

Smule's Ocarina (Wang 2009) is an imitation of the ocarina wind instrument that runs on the iPhone. The player blows into the microphone, which estimates the degree of pressure, and presses zones on the touch screen corresponding to the holes on the original instrument. However, in addition to these traditional aspects, tilting the phone allows for modification of the timbre (Figure 2.8).

An *augmented instrument* is an extension to a traditional acoustic instrument (Widmer et al. 2007). Typically, extra capabilities are provided through the addition of sensors that create additional sounds or control effects applied to the acoustic sound (Miranda and Wanderley 2006). For example, McPherson's (2010) Magnetic Resonator Piano uses the Moog Piano Bar (2005) to optically senses the position of each key on a grand piano. This input is then mapped to an array of magnetic transducers installed above the piano strings that individually vibrate each key's strings in response to the position of each key.

Finally, a number of notable commercial DMIs created in recent years are actually controllers designed to control arbitrary software 'instruments' running on a computer. Sometimes they are bundled with optional software but often with the assumption that many musicians are expected to advance from this when ready. Examples include the Eigenharp (2010) and the Karlax (2010).

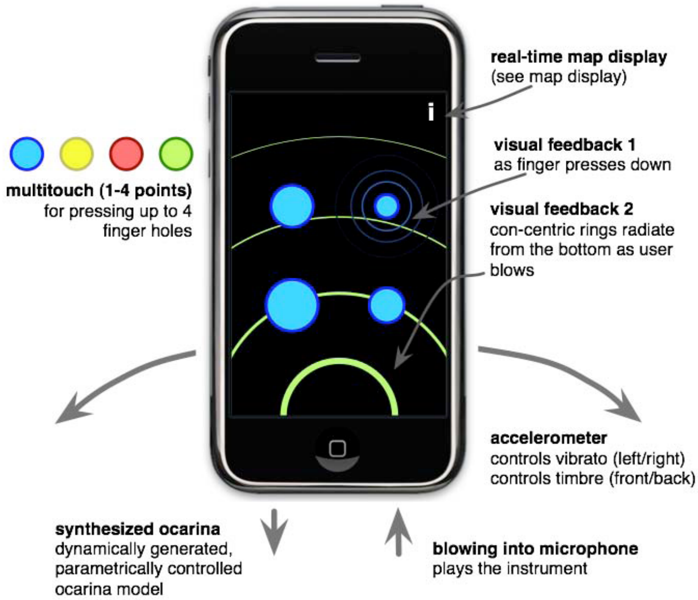


Figure 2.8: Design schematic for Smule's Ocarina. Reproduced with permission from Wang (2009).

Creating DMIs with character

A lesson from DMI research has been that increasing the capabilities of a system does not always improve it. Tanaka (2006) argues that it is the idiomatic characteristics of instruments—effectively their own unique limitations—that give them their *voice*. He illustrates this point with reference to Sensorband, a three person ensemble in which he performs. The three musicians all perform with unique and custom made DMIs that would be unknown to an audience. However, each has a unique way of articulating sounds allowing a listener to gradually unravel the voices making up the sound of the ensemble. In a similar vein, Magnusson (2010) proposes that with the boundless possibilities of modern synthesis software, we need to think less in terms of *affordances*—what an instrument permits its player to do—and instead focus on which *constraints* we need to impose.

The potential of constrained instruments to provoke original and creative musical styles was demonstrated in a study by Gurevich et al. (2010). They provided nine music students with a highly constrained instrument and gave them a week to prepare a solo performance, which was videoed and analysed following a formal coding process. The instrument was simply a box with a button that caused it to output a tone of fixed pitch and amplitude, some holes allowing the sound to escape and an on/off switch. However, a diversity of different performance techniques arose including tapping the box, filtering the sound by covering the speaker holes and creating timbral effects through operating the power switch whilst the button was pressed.

The dominance of the DMI paradigm

The Digital Musical Instrument has become the dominant paradigm within NIME research for systems that translate human action into sound for the benefit of an audience (Gurevich and Fyans 2011) including, for example, performances conducted entirely using a laptop (e.g. Zadel and Scavone 2006). However, considering all computer-based performance systems as drop-in replacements for musical instruments can place unnecessary constraints on the performance possibilities may arise from NIME research (Gurevich and Fyans 2011). Magnusson (2005), in the discussion of his and Hurtado's *ixi* software point out that the instrument paradigm was useful, but they outgrew it eventually. In the other direction, consideringeverything as an instrument risks overlooking what makes *instruments* different from *tools* (Tanaka 2006; Van Nort 2009). We will return to both of these arguments in Chapter 3.

Composed instruments

The *composed instrument* is defined by both Wanderley et al. (1998) and Fiebrink et al. (2010) as an instrument where the gestural controller is independent form the sound synthesis model. This is identical to our definition of a DMI (Section 2.2.2). However, Schnell and Battier (2002) refines this notion in a manner that we might expect from the name, as an instrument embodying some kind of precomposed musical work.

The term recognises that within NIME, the distinction between composition and instrument is somewhat blurred (Magnusson 2009; Drummond 2009). However, it is also an acknowledgement that prior to performing or improvising with a DMI, a degree of preparation may take place, such as defining mappings or loading samples (Dudas 2010). Such preparation is not exclusive to DMIs. Cage (1973) wrote for prepared piano and one may program presets into a synthesizer. However, as we will see in Section 2.2.7, the mapping of a DMI is fundamental in defining not only how it sounds but the types of musical interactions it lends itself towards.

Hahn and Bahn (2002) describe a composed instrument as a cohesive combination of controller, sounds, processing methods and display systems. Their work *Pikapika*, a performance where a wearable instrument was tailored around Hahn (who is a trained dancer), demonstrates the idea. The instrument was developed alongside the character of the show, with choreography, sound and interaction design emerging throughout the process with input from the experience of the dancer.

In Section 3.3.1, we will consider the relationship between composing instruments and composing music to extend this concept of a composed instrument.

2.2.3 Interactive Music Systems (IMSs)

We define an *Interactive Music System* (IMS) as a system that responds with music to input from a non-expert human participant without requiring assistance from a non-participatory human (e.g. an artist working behind the scenes). Our focus will be on systems that do so by use of a computer.

The two key types of IMS that are most suited to our formulation of interactive music (see Section 1.2) are installations and applications. We discuss both of these here—other relevant IMSs will then be covered in Sections 2.2.4 and 2.2.5.

Interactive sound installations

Interactive sound installations are a form of interactive art installed in a public space where visitors are able to influence what is heard. A key design consideration for this context is that few assumptions may be made about the audience, especially with regards to their experience interacting with similar systems (Dipper 2009). Visitors may not even be aware an installation is interactive. In this case, they are unlikely to engage with the system unless a perceptual feedback loop is quickly established (Schacher 2009). However, whilst it may be easy to get attention, it can be a challenge to provide longer 'nourishing' experiences (Machover 2002).

With interactive installations, the creator typically relinquishes control over timing aspects of their work (Klein 2009). Unlike, for example, on a smartphone app (or a lab-based evaluation), potential participants are likely to come and go individually at arbitrary times.

We saw Rokeby's *Very Nervous System* in Section 2.1.5 and Fels's *Iamascope* in Section 2.1.6 as examples of interactive sounds installations. Another example is Tanaka and Bongers's (2001) *Global String*, where a virtual musical string stretches across two distant locations with each end being physically realised as a steel cable within a public space. Visitors could interact through vibrating the string, which provided input to a synthesiser at both locations, as well as being actuated into physical vibrations on the string in the opposing space.

Further examples include Livingstone and Miranda's (2004) reactive space that responds to visitors' gestures as detected via video and *PlaySoundGround* (St. Clair and Leitman 2009), an installation where accelerometers were installed onto playground furniture to drive synthesiser parameters.

Interactive sound applications

Interactive sound applications might be considered similar to interactive sound installations except transplanted from a public space onto a personal computing device such as a laptop, mobile phone or games console. This leads to a number of distinctions with installations. The potential participant is less likely to ‘chance upon’ a work. There is less scope for learning through observing other participants. In addition, the application has the potential to remember its state and be aware when the participant is ready to interact, allowing it to personalise its behaviour to the user over time.

Interactive sound applications may also be referred to as sound or musical ‘toys’ (Robson 2001). However, we will avoid this term as it may be used in a derogatory manner to imply something is lacking in musical potential (e.g. Jordà 2004).

An example in this category is Block Jam (Newton-Dunn et al. 2003), an interface consisting of 26 physical blocks that magnetically connect. Sensors within the blocks detect the arrangement and use this configuration to drive a sequencer following the metaphor of bouncing virtual ‘cue balls’ through the connections.

Another example is Electroplankton designed by Iwai (2005) for the Nintendo DS. The user interacts with a set of animated marine animals, who each respond in a different way though animation and sound. Although it is created on what is predominantly a gaming platform, there are no explicit goals. Pichlmair and Kayali (2007) call this Active Score Music, where the artist has created an interface to a musical process for a user to explore.

With an application such as Electroplankton where user input may be reliably turned into sound, we may ask why it is not considered as a DMI. The potential diversity of outputs is often cited, which we will discuss in more detail in Section 2.2.11: in this sense we may think of Electoplankton as having a fixed number of pieces. However, in Chapter 6, I will propose creative agency as a more suitable means of answering this question.

2.2.4 Other performance paradigms

Digital Music Systems (DMSs) blur traditional distinctions between performer, composer, instrument and audience (Drummond 2009). Beyond DMIs and interactive sound installations a variety of different paradigms have been proposed. Whilst we will try to avoid where possible debates over the meaning of words, a key reason for establishing these paradigms is to facilitate communication of what kind of problem each system is attempting to solve and consequently which lessons might be applicable in different scenarios. Therefore, we will review here some of those that are most relevant. However, it is important to bear in mind that such paradigms are *descriptive* rather than *prescriptive* having been based on observations and generalisations rather than a requirement analysis (Linson 2011).

In this section, we review performance paradigms where a DMS is used by an expert performer for a non-participatory audience. Paradigms where DMSs are used without an expert performer follow in the subsequent section.

Interactive Composition, Intelligent Musical Instruments

Interactive composing is a term introduced by Chadabe in 1982 as 'a performance process wherein a performer shares control of the music by interacting with a musical instrument' (Chadabe 1997, p. 293). He describes it as a two stage process of creating a *interactive composing system* and then simultaneously performing and composing by interacting with that system. The requirements on the system are to function within a performance automatically and in real time to

- *Interpret a performer's actions as partial controls for the music,*
- *Generate controls for those aspects of the music not controlled by the performer,*
- *Direct the synthesizer in generating sounds* (Chadabe 1984, p. 23)

An *intelligent musical instrument*, a term also used by Chadabe to describe an interactive composing system, which he reports was introduced earlier by Laurie Spiegel and Max Mathews around 1973 as an instrument that analysed the performer's expressive intention and then automatically created the music (Chadabe 1997).

Mathews, reasoning that the notes of a piece were predetermined but a key source of difficulty, created the Radio Baton and Conductor Program as a sort of 'expressive sequencer,' allowing a listener to be more active and conduct their own interpretation of a piece. As Chadabe describes, it shares 'control of the music with the user, thus compensating for the user's lack of previously acquired skill' (Chadabe 1985).

Spiegel describes *intelligent instruments* as letting people play music on a compositional level.'[6] Keeping accessibility as a motivating factor, she created Music Mouse (Spiegel 1988), which used the two-dimensional coordinates of the mouse as a direct input to two on-screen keyboards. Further control from the computer keyboard altered a range of settings, which were then used in the automatic creation of additional musical material.

Thus we might be able to see interactive composing systems and intelligent instruments as two sides of the same desire to automate some aspects of the music making process, with the former being used in the conventional performance context and the latter in the informal improvising or home entertainment scenario. Chadabe sees a sense of performance-time unpredictability from the system as a crucial aspect of interactive composition as it allows mutual influence between performer and system. After all, guitar frets automate some aspect of the musical creation process and surely influence the actions of the performer but a guitar is not an interactive composition.

However, a number of questions remain unanswered. How far does the performer of an interactive composing system learn to use it? If a system is designed to provoke and behave in unexpected ways, does rehearsing with it to become more familiar with its behaviour make it less interactive? 'Unpredictable' is often used in a computer science sense, meaning non-deterministic behaviour—that is, a program in which providing identical inputs will at some stage give distinct outputs. But any musical systems receiving continuous input data is unlikely to receive an identical input twice. Even when 'unpredictable' is understood more subjectively as 'unexpected', to whom is

[6]Personal communication between Spiegel and Chadabe as quoted in Chadabe (1997, p. 334).

it unexpected? The audience? The performer? The creator of the system?

Finally, Chadabe (2004) describes an *interactive instrument* as a dynamic musical system with an identity of its own and that can 'seem to think for itself'. This requirement is familiar from our discussion in Section 2.1.3 of the many meanings of the word 'interactive'. The framework we describe in Chapter 6 will provide a more specific grasp of what it means for an instrument to 'seem to think for itself'.

Rowe's Interactive Music Systems

Rowe defined an *interactive music system* as a system 'whose behaviour changes in response to musical input'[7] (Rowe 1993, p. 1). Although not explicit in this definition, Rowe characterises computer music systems as those that listen to, and in turn respond to, a performer with some kind of originality in their response. They become an active participant in a musical performance occupying the same role as a human musician might.

Rowe's definition has been criticised for being too restrictive in considering only music as input (Jordà 2005) or for limiting its scope to fit within established performance practices (Paine 2002). However, Drummond (2009) points out that at the time it was written, the vast majority of music software and controllers were MIDI based. Their criticism is perhaps more of the all encompassing nature of the term *interactive music system* rather than the merit of such systems.

Rowe identifies a number of dimensions along which to classify interactive music systems.

The first classifies between *score-driven* and *performance-driven systems.* Rowe's conception of a score follows traditional ideas of collections of notes organised by pitch and time.

Score-driven systems make use of fragments of scores both in terms of what

[7] Rowe actually defines an *interactive computer music system* although the term *interactive music system* is used throughout his book and in his title so we assume he means the two to be synonymous.

they are listening for and how they respond.

Performance-driven systems, on the other hand, tend not to anticipate a particular score and listen out for more perceptual parameters such as density and regularity.

The second categorises the method used by the system to generate an output.

Sequenced methods make use of prerecorded fragments of music with perhaps some slight modifications of rhythm or dynamics.

Transformative methods also rely on musical material but this may be heard live rather than stored and transformed in some way so that the output may not be recognisable from the source.

Generative methods have only elementary source material such as scored scales or sets.

The third dimension distinguishes between *instrument* and *player* paradigms and is concerned with whether the role is assumed to fill the role of an instrument (resulting in a solo if used by a single performer) or a type of artificial player (resulting in more of a duet if used used by a single performer).

Jordà (2005) criticises these first two dimensions as being highly correlated. After all, a score-driven system is surely relying on sequenced methods. We may perhaps address this criticism by reinterpreting the first dimension as the range of *input* to which the system has been intended to respond and the second as how far the performer may predict the *output*. Alternatively, we may consider the first as relating more to the greater compositional structure and the second to more microscopic aspects.

An example of an interactive music system is Pachet's (2003) *Continuator*. The *Continuator* responds in real-time to MIDI input by generating a musically interesting but stylistically consistent response. The style model may be learned online as the user interacts with the system or a previously recorded model may be used. Interactions with the *Continuator* tend to follow a pattern of call and response. It fits neatly into Rowe's classification system as a performance-driven, generative system within the player paradigm. It also satisfies some of the more tricky definitions of *interactive*: interacting with it is conversational requiring perception, memory and cognition by both parties (Bongers 2000). It is capable of changing and evolving in response to these interactions (Paine 2002). However, Pachet (2003) points out that the stylistic models can only 'fool' the listener on a short scale. Responsibility remains with the performer to establish structure within a piece.

Another system that aligns with Rowe's definition is *Swarm Music* (Blackwell and Bentley 2002; Blackwell and Young 2004). Blackwell et al. draw on analogous concepts between group improvisation and the dynamics of a particle swarm such as *emergence* as structure arises in an unplanned manner from individual musicians and *attraction and repulsion.* Examples of this latter concept include the notes in a chord that may be seen as avoiding being too close to each other (though not too much) and being attracted to consonant intervals. We may also think of musicians responding to each other's motifs (attraction) but developing them into something of their own rather than purely repeating them (repulsion). Blackwell et al.'s system interacts with a performer through

1. Parsing an audio or MIDI input into events,
2. Mapping them into attractors within a multidimensional vector space with different dimensions representing different musical parameters,
3. Applying the swarm update function to a (pervasive) set of particles within the space,
4. Mapping the set of particles back into audio events.

In terms of Rowe's classification, this system would be performance-driven and within the player paradigm. Arguably, it is both a transformative and generative system being driven ultimately by a chaotic generative system but providing responses that are directly based on previously heard material.

As mentioned in Section 2.2.3, we are using the term *interactive music system* in a more broad sense than Rowe. Where we wish to refer to Rowe's definition, we will explicitly write Rowe's Interactive Music System.

Interactive dance / Trans-domain mapping

Pikapika, discussed in the Section 2.2.2, is also an example of an interactive dance performance. *Interactive dance* refers to the use of audio or visual systems within a dance performance that are influenced by the actions of the dancer (Bevilacqua et al. 2011). Gestures are typically sensed through

- motion capture[8](e.g. Bevilacqua et al. 2001; James et al. 2006),
- off-stage cameras (e.g. Camurri et al. 2000, 2003; Ng 2002a)—often using a computer vision toolkit such as EyesWeb (Camurri et al. 2004) or Music-via-Motion (Ng 2004),
- wireless accelerometers attached to the dancer (e.g. Park et al. 2006; Morales-Manzanares and Morales 1997; Aylward and Paradiso 2006),
- a controller held by the dancer (e.g. Murray-Browne et al. 2010).

Creating an interactive music system that responds to dance presents a unique set of challenges. Dance is a predominantly visual medium and it will often be desired for hardware to be hidden. It is also common to avoid using hardware that imposes physical constraints, both in terms of bodily movement as well as the dancer's position within the performance space.

Of greater interest to us, however, is the question of how the dancer exerts influence over the music. Typically we would not want to limit our potential performers to dancers who are also musicians and, regardless of

[8]Motion capture refers to systems such as those produced by Vicon (`www.vicon.com`) that sense the skeletal position of a person through attaching reflective markers to their limbs and inferring their 3D position using an array of infrared cameras. Such systems are accurate but expensive and not very portable (Geroch 2004).

whether or not they are, there is not usually the time to learn a new musical instrument for a single show. However, our dancer most likely is a highly skilled performer with a mastery of expressive movement. For these reasons, systems that provide low level control tend to be inappropriate (Siegel 2009). Not only will the dancer be ill-equipped to express themselves musically but the detailed gestures required to exert this control are likely to prevent them from fully expressing themselves through movement. In the language of gesture (see Section 2.2.6), we might consider dance as a medium entirely of ancillary gestures that risks being disrupted through the introduction of technical gestures. Therefore, with interactive dance it is the task of the designer to adapt to the gestural language of the dancer, rather than the dancer the gestural language of the designer. However, low level mappings may make a brief appearance. We see this, for example, in the interactive dance performance *CoIN* (Ng 2002b) which included a section where absolute position of the hands of dancers were mapped to pitch.. Although, as Ng reports, it becomes quickly tiresome, simple low level mappings were used early in the performance to demonstrate to the audience that movement and sound were related.

Thus, the aim with interactive dance systems is typically to provide the dancer with high level control over the *expressive* aspects of the music, providing an interface that is both intuitive and lets them feel they are in control (Morales-Manzanares et al. 2001; Siegel 2009). This requirement is similar to Spiegel's intelligent instruments (Section 2.2.4). However, it is worth remembering that unlike Spiegel's users, a dancer is a *performer*. They are trained to express themselves and experienced in structuring an entire performance. Ng (2002a) describes this process of translating activities from one creative domain to another as *trans-domain mapping*.

The introduction of material in *CoIN* demonstrates another important difference between interactive dance performance and a concert. In a concert, the audience arrives with the expectation that the performer will be creating the music. This cannot be taken for granted in a dance performance (Siegel

2009). We will consider this issue in more depth in Section 3.3.2.

Other performance paradigms

Finally, we will briefly outline a number of other performance paradigms that are not a direct part of the argument of this thesis but are relevant to understand its context.

Tape music is simply the listening of an audio recording in a concert environment (Rowe 1999). It is common to mix tape music with live instrumentalists. This practice remains common in many modern gig scenarios where a band play on top of a sequenced backing track (Robertson 2009). Examples performed in a more classical context include Steve Reich's *Different Trains* (1988) for (unamplified) string quartet and tape. Rowe (1993, 1999) sees the mixing of performers and tape as an expression of the desire to include human musicianship within computer music but identifies a number of limitations. For example, the performance is restricted to a fixed tempo and both instrumentalist and tape cannot make a simultaneous attack after a considerable period of silence unless audio or visual feedback is provided to the performer.

An important lesson to draw from tape music is that the concert performance scenario has more to offer than simply providing a space for a musician to perform. As Rowe argues,

> the communal experience of hearing tape music in a large space through superlative sound equipment is a valid and exhilarating aesthetic venture in its own right. (Rowe 1999, p. 86)

We will explore the nature of concert experience in more detail in Chapter 6.

An *accompaniment system* is defined by Roads et al. (1996, p. 681) as a system that 'plays along with a human musician, following the performance of a score by the human, and playing its own score simultaneously.' These are often created with the clear goal of providing more flexibility in timing than a traditional tape recording when musicians are performing with a backing track (Robertson 2009).

Similar in definition, Jordà (2005) argues that a system that only allows for a predefined piece to be played in a performance is a *score follower*, a definition that itself relies on the definition of the term 'piece'. Under this interpretation, we may also include systems that are not designed necessarily for musicians. For example Castellano et al. (2007a,b) created a system that renders a performance of a fixed piece of music with the body position and movement of a participant controlling parameters such as tempo and sound level. More detail on score following systems may be found in Stark (2011).

Finally, *live coding* is a type of musical performance that involves creating music through programming during the performance (Collins et al. 2004). As a result, it permits a high level of abstract process-oriented control (Brown and Sorensen 2009). An interesting aspect of live coding is that the code is usually made visible to the audience as well as the performer (Magnusson 2011). It is also common for performers to provide comments within their code to assist the audience's understanding of it (Fyans et al. 2010). Although live coding does not form a part of the argument of this thesis, it is relevant to Chapters 3 and 6 when we consider how an audience forms an understanding of performed musical interaction—and what effect this has on their ability to appreciate what they are hearing.

2.2.5 Other participatory paradigms

In Section 2.2.3, we stated that interactive sound installations and applications are the most suitable domain for our formulation of interactive music. In this section we review a number of other relevant participatory paradigms that fall within our definition of an IMS.

Music games

Music games refer to videogames that embody a strong musical focus. Recent examples include *Guitar Hero* (Harmonix Music Systems 2005) and *Rock Band* (Harmonix Music Systems 2007). Both of these games are often referred to as *rhythm games* as the objective of the game is to push buttons in time to the music (as indicated by a scrolling piano-roll style score) (Richardson and Kim 2011). As musically-related as these tasks may get, there remains an objective evaluation criteria determining whether a player performed correctly or not. Other games may provide music as a reward for progress (e.g. Saint and Stallwood 2008).

Rosenstock (2010) attempted to incorporate more improvisatory (and hence less objectively evaluated) aspects in the game *iGotBand* where players could collect 'fans' by playing sequences of notes that contain a fan's requested notes. However, Rosenstock and the other developers struggled to arrive at a meaningful condition for winning the game without completely disregarding all improvisatory aspects.

Whilst interesting in their own right, music games are crucially different than interactive music. Objective and explicit goals are established, with the player's input scored against an 'ideal'. Where an explicitly stated and objective goal has been established, a play activity has become a game (Sykes 2006). Musical ideas may inspire these goals, and music may serve as the reward, but ultimately the motivation is not to explore and create but to win.

Reactive music

Reactive music is a term popularised by the creators of the software RjDj (2008). We will use the term to describe music systems in which non-musical control influences what is heard such as the sound of the user's environment (e.g. RjDj 2008; Vawter 2006) or incidental gestures that were not performed primarily to be sensed by the system such as walking (e.g. Inception 2010) or subconscious responses to the music (e.g. Tanaka et al. 2005). These features are often facilitated by running custom software on a phone. Reactive music compositions retain a definite author but they are typically constructed to respond continuously to input (Nordgren 2009). Most notable is RjDj (2008), an iPhone app that implements PureData (Puckette 1996) as well as providing a platform for composers to distribute reactive music scenes and users to record and share their realisations (Nordgren 2009).

Creators of reactive music are often motivated by a desire to make personal music listening less isolating—through connecting with the immediate surroundings (e.g. Vawter 2006), collaborating with others (e.g. Tanaka 2007) or sharing recordings (e.g. RjDj 2008)—as well as providing listeners the opportunity to be creatively involved in their music (e.g. Gaye et al. 2010; Tanaka 2007).

Tanaka's (2004a; 2004b) Malleable Mobile is a portable handheld system that uses involuntary body gestures (grip pressure and acceleration) as well as deliberate interaction with a touchscreen to drive a personal generative music stream, which is then mixed with other users who are in range. Musical works for Malleable Mobile are in an open form that combines elemental modules such as rhythm generators, time stretching algorithms, sequencers and samplers. Tanaka highlights the need to allow the participant to distinguish their own personal music stream from that of remote partners.

Another example of a reactive music system is Sensory Threads (Marshall et al. 2010), in which groups of four individual urban explorers are each given a separate sensor which connects through bluetooth to a single portable computer that generates a shared soundscape.

The defining difference between reactive music and our described interactive sound applications (Section 2.2.3) is the focus of the participant. With a reactive system, the participant interacts indirectly with the system via their environment, with or without intention (Figure 2.9).

Adaptive music

Adaptive music describes a videogame soundtrack that adapts to the state of a game (Clark 2007). The user interacts with the game but without the explicit goal of interacting with the music (Collins 2007). As a soundtrack, the music needs to remain contextually appropriate. An additional demand made of adaptive music is to allow musical material to be reused without

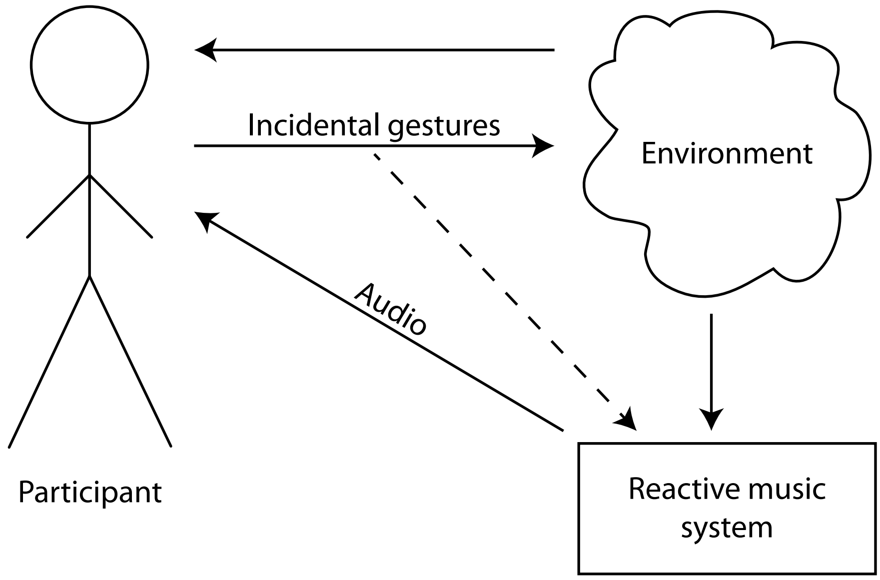


Figure 2.9: The flow of influence within a reactive music system. The participant interacts incidentally with the environment. The system then produces audio that is influenced by the environment and sometimes incidental gestures.

seeming overly repetitive. Many modern games may be played for up to 60 hours but do not carry nearly as much audio as this (Collins 2009).

The most common procedural techniques used to perform adaptive music are horizontal resequencing, where segments of a score are reordered, and vertical reorchestration, where different tracks of the score are remixed in realtime (McAlpine et al. 2009; Collins 2008).

Game composer Hannigan argues that a videogame player does not necessarily want to be told *what* to feel (Durrani 2009). Therefore, adaptive soundtracks may often be fairly ambient in how they sound, avoiding the overtly emotionally charged material familiar from film scores. For example, Nutt (2008) created a soundtrack system that simply places sound samples within the game space to be played without any synchronisation and just a few rules to prevent clashes.

As with reactive music, even though the user is directly interacting with a model that drives the sound, adaptive music is distinct from interactive as it is the state of the model rather than the music that is the focus of the user's attention (Figure 2.10). Whilst adaptive music may enhance the user's experience, it is not typically the primary output that motivates them to do so. Furthermore, distinguishing it from music games, reactive music does not typically serve a rewarding function for the attainment of goals.

Each of the above examples demonstrates IMSs where the participant interacts with a different system or environment in order to influence what they hear. However, these paradigms are not necessarily mutually exclusive; nor are the boundaries between them clear cut. Our descriptions have focused primarily on who is influencing whom and in what situation. The conceptual tools we develop Chapters 4 and 6 will provide further perspective on the effect this will have on a participant's subjective experience.

Any kind of Digital Music System (DMS) creates sonic output in response to a human input and there is a wealth of research into how design decisions

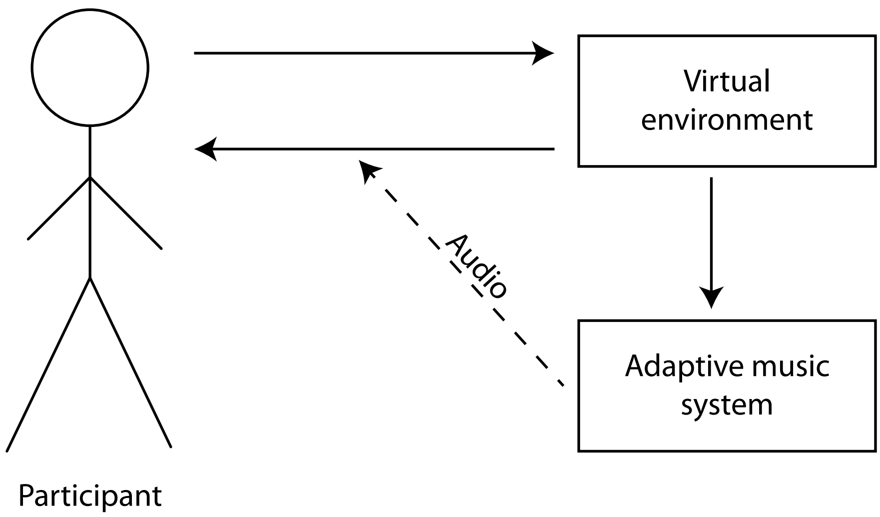


Figure 2.10: The flow of influence within an adaptive music system. The participant interacts with a virtual environment. The system produces audio that is influenced by the the state of the environment and forms a part of the overall feedback from the virtual environment.

affect the experience of both the user and an audience. Although much of this research has been carried out under the DMI paradigm, there are some key lessons that are relevant to our investigation. In the remainder of this section, we outline *gestures* and *mappings* as the key tools used to reason about DMSs, as well as a number of important issues to be considered when designing them.

2.2.6 Gesture

The word *gesture* typically describes any kind of musically meaningful human movement that may be sensed by an interactive system (Wanderley and Orio 2002; Malloch et al. 2006) or as any action produced by an instrumentalist during a performance (Wanderley and Depalle 2004).

Musical gestures may be classified as *manipulative* or *ancillary* (Miranda and Wanderley 2006). Manipulative gestures are those that affect the sound produced by an instrument. Ancillary are the remaining gestures musicians produce when they perform, such as a pianist swaying their upper torso (Wanderley et al. 2005). Ancillary gestures provide a means of further visual communication and have been shown to influence how the expression of a musical performance is interpreted by an audience (assuming they can see the performer) (Davidson 1993). Wanderley et al. (2005) found that seeing as well as hearing a clarinetist perform could heighten or dampen the tension perceived by a listener depending on the motion of the performer. There may also be interactions between what is heard and what is seen. For example, both auditory and visual combine to inform an audience of phrasing information (Vines et al. 2006). Of course, non-musical gestures also communicate affect (e.g. Sanghvi et al. 2011).

Cadoz and Wanderley (2000) argue that considering gesture solely as a one-way information channel from performer to instrument is simplistic. As well as means of communicating information (the *semiotic* function) and manipulating the environment (the *ergotic* function), gestures serve as a means of perception through tactile feedback (the *epistemic* function). These functions cannot be isolated: manipulation is often to facilitate perception whilst tactile perception may affect how gestures are formed.

How we move is intimately linked with expression and emotion through our embodied relationship with the world (Davies 2001). We also have intuitions about how sound and movement relate (Levitin et al. 2002). For example, it seems more natural for more energetic movements to create louder sounds (Hunt et al. 2003). Antle et al. describe these natural associations between abstract concepts and our physical relationship with the world as *embodied metaphors*. In a study, they interviewed a small number of dancers and choreographers to derive a set of embodied metaphors relating music and action. They then compared a mapping based around these metaphors with a non-metaphorical mapping in a study where inexperienced adults and children were given between four and six musical tasks to perform within an interactive sound installation. Participants were permitted up to ten minutes to practice each task. They found that although there was no significant difference in practice time, the metaphorical mapping significantly improved accuracy in all but one of the tasks (Antle et al. 2009). As well as carrying implications about how to make systems that are easy to learn, this suggests that these embodied sound/action relationships are somewhat invariant among individuals, although we should note that this was not a cross-cultural study.

2.2.7 The mapping model

The specification by which gesture results in sound is typically reasoned about in terms of information processing (e.g. Kvifte and Jensenius 2006) through a layered model, consisting of an input *controller*, an output sound *generator* and a *mapping* defining how the former affects the latter (Figure 2.11) (Drummond 2009).

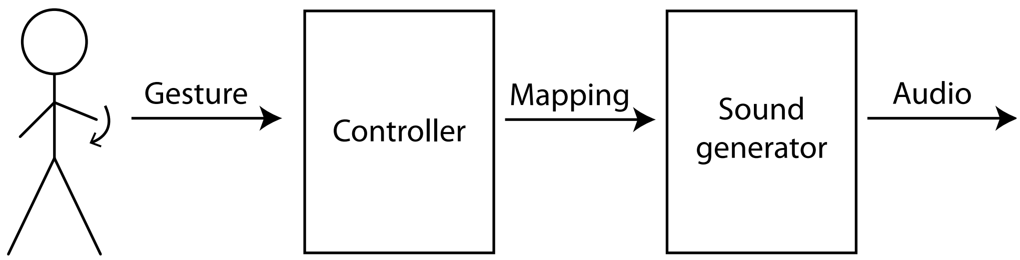


Figure 2.11: The mapping model. Gestures are sensed by a controller, the output of which then controls the sound generator in a manner specified by the mapping.

Controllers and generators

There are a diverse range of controllers used to measure some aspect of a gesture. Detail of controllers is not necessary for the argument of this PhD, however we provide here a brief overview of what types of interaction they make possible.

Controllers vary in accuracy, inconvenience, precision, expense, reliability (Fels and Lyons 2011). Different controllers lead to different types of input signals. A key difference that arises is between event-based (i.e. triggers) and continuous data (Stowell 2010). Continuous data will differ in terms of the shape of signal provided (Tanaka 2006). This is a feature of a sensor, but also of how it is used, for example to measure movement of the arms vs that of the fingers.

All of these different properties have differing musical strengths and this depends to a large part on people's differing means of communicating. Sensors that can detect grand sweeping gestures of the body may be suited for continuously controlling dynamics but may not be so suitable for controlling pitch, for example (Fels and Lyons 2011).

Some controllers are suitable for provide fixed input data (e.g. a ribbon pad). However, if there is not a clearly fixed origin against which to base measurements then relative data might be more appropriate (e.g. when using a camera or accelerometer). Non-contact controllers such as the theremin are poor for high resolution control due to their lack of direct physical interaction (Hunt and Hermann 2004).

A controller will suggest different gestures. Likewise, the gestures an instrument designer wants to capture will suggest different controllers.

Types of mapping

Mapping refers to how data from the controller influences the sound generator (Hunt et al. 2000). Typically, we think of simple signal processing of the data (e.g. blob detection of a video or smoothing of an accelerometer) as a part of the controller. The controller will then provide a number of outputs, which are used to determine the parameters of the generator.

When discussing mappings, the *input* to the mapping is the output of the controller and the *output* of the mapping is the input of the generator. Mappings are often represented as a graph of arrows indicating the flow of information (Figure 2.12), which sometimes leads to confusion. Mathematically,

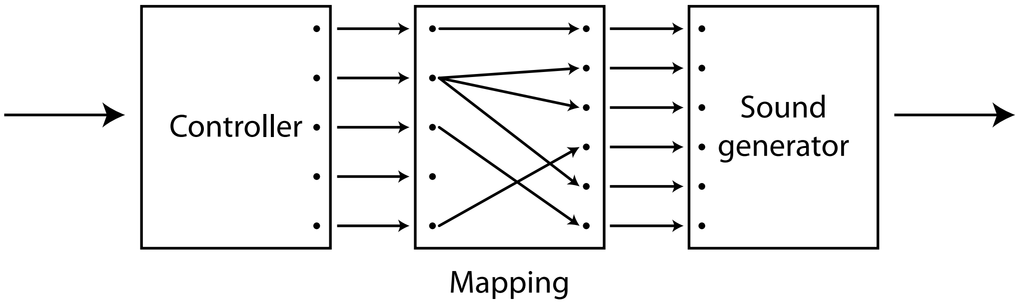


Figure 2.12: A mapping represented as a flow of information from outputs of the controller to inputs of the generator. Note that it is the mapping is defined over the vector space defined by the inputs and outputs rather than the components as might be suggested by this type of diagram.

a well-defined mapping relates elements of one set (our input) with elements of a another set (our output) such that every element of the former is mapped to a single element in the latter. Jordà (2005) argues that this makes the term *mapping* somewhat inappropriate as it neglects the possibility of two controller outputs being used together to control a single parameter. However, the *mapping* refers to how elements in the *vector space* defined by the inputs are mapped to the vector space defined by the outputs (Van Nort et al. 2004) and not, as may be implied by the diagrams used, how components of the input are paired with components of the output. This confusion arises because mappings are typically *reasoned about* in terms of which input components determine which output components. To support this line of reasoning we think of the *relation* between output parameters and input parameters. When we refer to the input*s* and output*s*, we are referring to the determinacy relationship between different components of the input and output. Therefore, diagrams showing arrows pointing from the components of a controller output to the components of an input indicate that a particular component on the output is determined by a specific subset of the components of the input space. They do not usually indicate *how* these subspaces are related, although this may play as crucial a role if not more so

(Van Nort and Wanderley 2006).

Mappings are often classified as one-to-one, many-to-one, one-to-many and, combining these last two, many-to-many (Hunt and Wanderley 2002). This refers to the dependency relationship between inputs and outputs. Whilst one-to-one may seem to offer the simplest means of controlling the generator, there is evidence that complex mappings involving many-to-many (possibly non-linear) relationships may be more intuitive and expressive. Hunt et al. (2003) conducted an experiment comparing a set of sliders mapped to a synthesiser in a one-to-one fashion with those mapped in a more complex fashion where, for example, volume was controlled by the speed of one slider meaning it had to be 'bowed'. They found that not only was the complex mapping found to be subjectively more expressive and easier but also, using a set of objective 'musical tasks', that participants' ability to reproduce specific outputs increased more quickly. One of the conclusions we may draw from the above-mentioned study by Hunt et al. (2003) is that the mapping that is simplest to describe explicitly is not necessarily the one that is easiest to control.

Jordà (2005) illustrates the complexity of the 'mappings' of acoustic instruments by considering the violin. The volume of a violin is dependent not only on the speed the strings are bowed, but the angle and pressure of the bow (among other things). Likewise, the pressure of the bow controls not only volume but also timbre. Most mappings map a moderate number of controller outputs to a much larger number of generator inputs so could be thought of as few-to-many (Jordà 2005).

As well as a controller lending itself to certain types of gesture, it also lends itself to certain types of mappings. Likewise, different generators will suggest different mappings. The mapping defines the limits of what is possible, both in terms of its theoretical limits but also in terms of what kind of physical action is necessary to produce a given output. The mapping also, in combination with the controller, determines the first sounds a newcomer will make with an instrument, thus defining 'normal playing' (Arfib et al. 2003).

Given a fixed controller and generator, different mappings will make it easier or hard to produce different outputs. As a result, different mappings will be suggestive of different kinds of actions as well as different generated sounds. In this sense the mapping defines to a large extent the *feel* of the instrument (Van Nort and Wanderley 2006).

Perceptual layers

Some have found it helpful when creating and reasoning about mappings to divide the mapping function into a composition of separate functions, with the intermediate spaces creating mapping *layers*.

Wanderley et al. (1998) used an inner layer of *abstract* parameters, such as fundamental pitch, vibrato, loudness. Their approach is somewhat motivated by a desire to provide a common set of 'universal' perceptual parameters than any controller might be expected to control and any generator might be expected to respond to such as timbre, loudness and pitch (Wanderley 2001a). This allows a notion of the choice of synthesis algorithm being 'transparent' to the user (Hunt and Wanderley 2002).

Other researchers have expanded on this model with extra abstract layers, sometimes providing alternative means of feedback from these middle layers (Figure 2.13). This leads to a layer of 'meaningful' gestural parameters that are independent of sound and layer of 'meaningful' synthesis parameters independent of gesture (Arfib et al. 2003; Hunt et al. 2003). In the end, however, these layers may become embedded within our notion of the controller or generator itself.

Dynamic mappings, nondeterminism and states

The first possibility we may notice unaddressed by the mapping model so far is nondeterminism. This is easily added by adding inputs from a random source. We may also conceive of a dynamic mapping that changes over time, and include time as an input component (Figure 2.14).

We may also consider a mapping that dynamically adapts to a user input

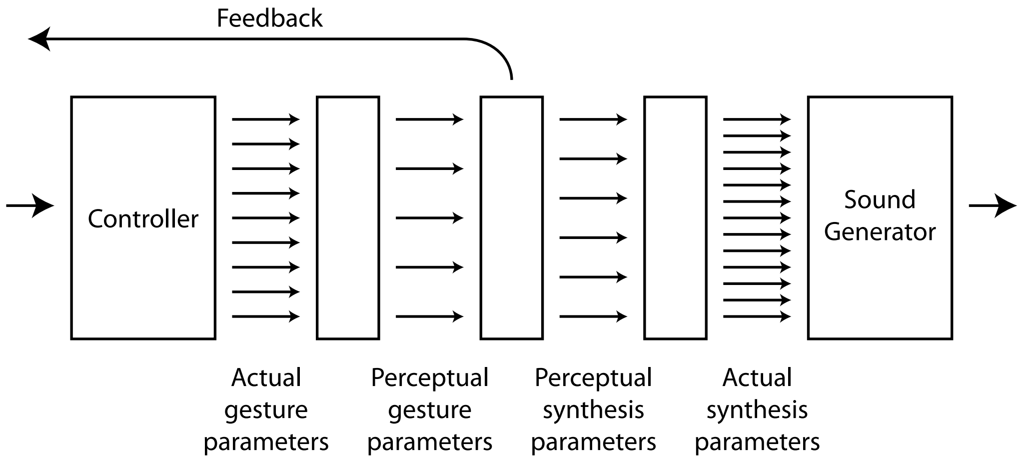


Figure 2.13: A mapping with perceptual layers. The overall mapping from controller to generator is defined by three smaller mappings shown here by unmarked rectangles. The innermost of these defines additional feedback for the player. Adapted from Arfib et al. (2003).

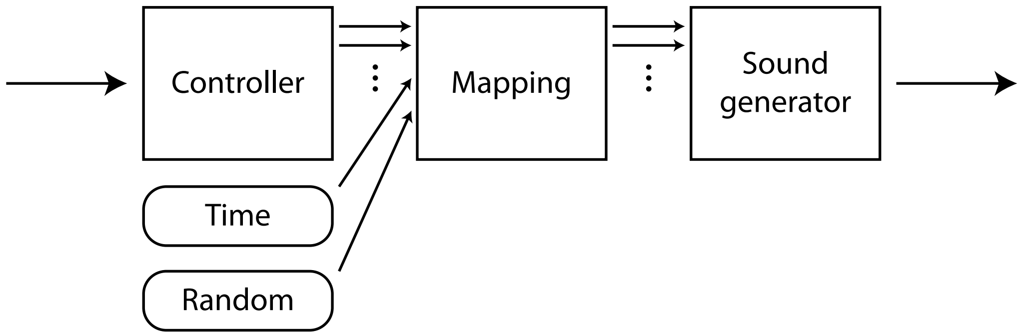


Figure 2.14: A dynamic non-deterministic mapping.

over time (Arfib et al. 2003). This requires us to revise our model to permit the system to have a *state*. Fels et al. (2002) describes these as *modal* mappings, where the state (or mode) of the system will influence how input is mapped to output (e.g. a synthesiser keyboard with different buttons allowing different patches to be selected). The state of the system may not only interpret how input causes output but also how input further affects the internal state. For example, any mouse and pointer based system relies on a model of where the pointer currently is to interpret how mouse movement determines the pointer's new position. Other systems will rely on more complex internal models (e.g. Magnusson 2005; Johnston et al. 2009). We may add this to the model (Figure 2.15). The model now looks similar to

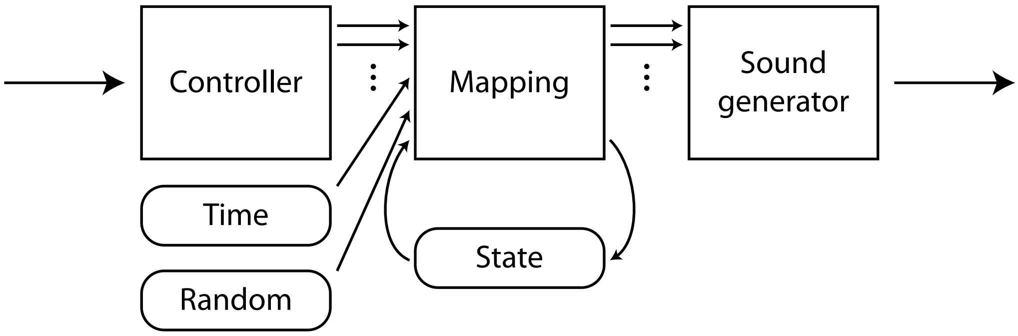


Figure 2.15: A dynamic non-deterministic adaptive mapping.

system 4 of Cornock and Edmonds's (1973) classification with the addition of a controller and sound generator (Section 2.1.2, see Figure 2.2d).

With modal systems, some gestures may have immediately perceptible effects whereas others may have long-term effects but no immediate ones. Camurri et al. (2004) describes this characteristic as the *strategy* of a mapping, ranging from *reactive* (immediately perceptible) to *dialogical* (not immediately perceptible but still determining the output). We saw this distinction applied with *Absolute_4.5* (Section 2.1.2).

Feedback

Feedback refers to information received by the user about the system they are interacting with. It closes the perceptual loop (Section 2.1.1) and allows the performer to understand the effect of their input to the system. In the context of DMI performance, this allows the performer to monitor the sound produced by the instrument and adjust their actions as necessary to create the sound they expect (Widmer et al. 2007). In other contexts, it also makes learning through exploration possible (e.g. with the Reactable (Jordà 2003)). As well as the sound produced by an instrument, feedback may be haptic (what is felt) and visual (Jordà 2003). Haptic feedback may then be *tactile*, the sense of touch felt by the skin, or *kinaesthetic*, the resistance felt by joints and muscles (also called proprioceptive feedback) (Bongers 1994).

Wanderley (2001a) classifies feedback in two dimensions.

Primary/secondary. Secondary feedback is the sound produced by the instrument. Primary feedback encompasses all other forms such as visual, auditory (e.g. the noises made by the keys on a clarinet) and haptic.

Active/passive. Passive feedback arises through the physical characteristics of the system (e.g. the clicking of a switch) whereas active arises from the output of the system.

The classifications are somewhat ambiguous when generalised beyond the DMI paradigm. For example, if the instrument is also driving a visualisation, would that be primary or secondary? However, we may perhaps interpret primary feedback as distinguishing that which is also received by the audience. This highlights some of the peculiarities of interactive music. Ordinarily, if our focus is on a single participant, one might argue that this distinction is somewhat irrelevant, or that all of the feedback is primary as it is all contributing towards that individual's experience. However, when bystanders are present and able to see a portion of the system's output, then the participant—intentionally or not—is in the role of a performer and primary and secondary feedback becomes relevant again.

Haptic feedback is often identified as an important aspect of acoustic instruments that DMIs often lack (e.g. Dobrian and Koppelman 2006). With acoustic instruments, the controller tends to be physically involved with the sound generation. This may provide tactile feedback in the form of acoustic vibrations as well as the feel of the physical shape of the instrument (Hunt and Hermann 2004). Some have identified this lack of vibrational feedback as a barrier to an instrumentalist forming an intimate relationship with DMIs and attempted to address this problem through adding force feedback to DMI controllers (Chafe 1993; Marshall and Wanderley 2006).[9] In terms of the physical shape of DMI controllers, most do indeed offer the same feedback as acoustic instruments (hands-free systems such as the theremin being exceptions). However, Jordà (2005) identifies the slightly more subtle issue. The physical form of the 'controls' on acoustic instruments typically arises directly from the acoustic process being controlled. Their shape is therefore communicative of their role in a way that those on generalised DMI controllers are not.

Whilst primary feedback is typically the means by which a musician evaluates the music they are producing, it will only ever be heard after it has already occurred. It is therefore difficult to perform with an instrument without any secondary feedback (Dobrian and Koppelman 2006). For acoustic instruments, Jordà (2005) identifies haptic as most important after auditory with visual being less important although he does not see this as a reason that DMIs should not make full use of visuals as a means to communicate information back to the player (Jordà 2003).[10]

This potential for visual feedback he demonstrated in the Reactable (Jordà et al. 2005, see also Section 2.2.2), Whilst perhaps essential for systems such as these where exerting control requires knowledge of the state of an internal model it may also be shared with the audience to aid their understanding of how the music is being created (Jordà 2003). With collaborative instruments, visual feedback is useful in allowing players to identify their own contribution (Blaine and Perkis 2000) and its communicative capacity also allows it to aid learning where input is being interpreted in a manner not be immediately apparent (e.g. Lyons et al. 2003). However, it is worth considering that there may be reasons that, once learnt, traditional instruments tend not to rely

[9] As a pianist who regularly uses an electric piano with headphones, I do not particularly identify with this need.

[10] I would argue (again perhaps from a pianist's perspective) that, once an instrument has been learnt, performing without any sense of touch would be far more difficult than performing without being able to hear.

on visual feedback. For example, communication through both gesture and eye contact plays a fundamental role in conventional ensemble performance (Seddon and Biasutti 2009) and may be impeded if the performers cannot look away from their instruments.

Visual feedback may also be useful within an installation context. For example, Kobori et al.'s (2006) interactive installation *LINE* involves a simple mapping between gesture and a coloured light beams, which then drives more complex generative sounds. However, in an installation such as this, the visuals form part of the primary feedback and may detract focus from what is heard. Castellano et al. (2007b) interviewed users of an installation where their movement and posture controlled both the a projection and the rendering of a piece of music. They found that whilst nearly all users felt control over the visuals, less than half did over the music.

Model limitations

The mapping model promotes a certain approach creating DMIs. We may identify two main drawbacks to this approach.

Firstly, the mapping model divides instruments into three separate components to be studied independently: controllers, generators and mappings to join them together. Jordà (2005) criticises this reductionist approach. He points out that much research has focused on controlling instantly perceptible synthesis parameters rather than more complex musical control processes, and that these results cannot necessarily be abstracted away from the controllers used in a given study. When things become more complex, creating a sophisticated controller requires knowledge of how the generator works and a more holistic approach is needed (Jordà 2005). The mapping plays a pivotal role in how the sound is controlled and should not be thought of simply as the 'glue' connecting controller to generator (Van Nort 2009).

Secondly, the model arrives with many preconceptions originating from conventional instruments. Everything is reasoned in terms of how input is transformed into output. Information flows from the performer into the instrument into the sound output, with any influence being returned back to the performer being feedback to aid the performer it achieving their musical goals. By contrast, Chadabe (2002) considers controlling interactive music systems in a manner similar to 'fly-by-wire' aviation where a computer controls a plane through interpreting high level instructions from the pilot in the context of the current state of the plane. As we saw above, allowing the system to be unpredictable, to create ideas or even to automate simple musical processes requires increasingly more complex modifications to the mapping model. We soon find ourselves in the problem we identified with Cornock and Edmonds's model (Section 2.1.2), where a system is reasoned about in terms of how it *works* rather than how it is *perceived*.

Despite the limitations identified, the mapping model remains useful. Any system that establishes a perceptual feedback loop requires its use to create at least *something* of a mental model of how action leads to results. The problem arises when a system is considered *entirely* in terms of an input-mapping-output model.

The concept of a mapping will be essential in Chapter 4 when we develop our model of how IMSs are explored. Over the course of this thesis, we argue the position that mappings as outlined above are less useful in terms of system specifications and more so when considered in terms of mental models.

There is no common set of criteria by which to evaluate musical interfaces (Paradiso and O'Modhrain 2003). Jordà (2005) points out that there can be no complete criteria by which one instrument could be described as 'better' than another as any evaluation will rely on context and personal taste. However, he argues that there are a number of common criteria by which a designer might evaluate their success, what Wessel and Wright (2002) describe as *features desired for all instruments*.

For the remainder of this section, we will review some of these desirable features. Note that most of these have usually been presented as criteria for *instruments*. This criteria should not be read as one that we will necessarily adopt. However, the concepts, as well as the reasoning behind them, are invaluable as a means to consider IMSs in general.

2.2.8 Expressivity and the emotional pipeline

Expression is commonly used to mean the communication of meaning or feeling (e.g. Fels et al. 2002). In a musical context, it is a term very much bound up with the Western concert tradition. The Oxford Dictionary of Music (Kennedy 1996) defines it as

> That part of a composer's music such as subtle nuances of dynamics which he has no full means of committing to paper and must leave to the artistic perception and insight of the executant. . . .

This definition is similar to that used by Mion et al. (2010), who see it as providing material beyond the rational content of a message that may help its interpretation or make it more pleasant. This leads to the notion of expression as the emotional aspects of a performance that are beyond what is scored. A good performer interprets the emotional intentions of a composer, 'encodes' them in their performance through subtle rhythmic, tempo and dynamic changes. These are heard and then 'decoded' by the listener who then feels the appropriate emotion (Poepel 2005). This line of reasoning leads to the 'emotional pipeline' whereby music is seen as some kind of carrier signal used to transmit emotions between people.

The limitation of this perspective to scored music is recognised by Dobrian and Koppelman (2006), who expands upon this definition to include, for example, improvisatory music as 'those characteristics of the live performance that enhance [the effective conveyance of meaning or feeling] but are not contained in materials on a notated page or pre-programmed algorithm'. This is similar to what Jordà (2005, p. 229) presents as a 'naive' definition:

‘expression is a matter of finding personal ways to transmit one’s own ideas (not reproducing dated romantic clichés or templates.’ An ‘expressive instrument’ is thus one that allows its performer, given enough skill, to transmit their feelings through music. The instrument does not have anything to express (because it is a machine (Jordà 2005)). (Nor, by this argument, should the instrument *be* an expression of its designer but we will return to this topic in Section 2.2.11.)

Gurevich and Treviño (2007) take exception to this communicative model of musical performance as an hegemonic demand that new musical interfaces fit within the classical concert tradition. They, alongside others such as Small (1998) and Clarke (2006), take a more *ecological* approach to music-making, focusing on the relationships between composers, performers, listeners as they exist within specific contexts. They criticise the requirement of music to have predetermined expressive content, drawing parallels with the visual arts where, for example, an artist may create a work in response to a material rather than an emotional complex (Gurevich and Treviño 2007).

Thus, ‘expressivity’ has a diversity of meanings. Within the Western score-based tradition, it has a particular specific interpretation, and Gurevich and Treviño (2007) point out that the proposed ecological interpretation does not necessarily find it at fault. Rather, it acknowledges that it is one of many potential kinds of musical practice for which a musical instrument might be created. Likewise, the emotional pipeline model is useful, for example in the context of automatically rendering MIDI to audio (Livingstone et al. 2007). But as we will explore in Chapter 3, the narrow definitions presented above unnecessarily restrict some of the performative routes made possible by NIME research.

Where necessary, we will refer to *classical expressivity* and *ecological expressivity* to clarify whether we are thinking in terms of Dobrian and Koppelman’s or Gurevich and Treviño’s interpretation respectively. Note that this should not be taken to mean that Dobrian and Koppelman’s interpretation is limited to classical concert performance. It is also common to other performance practices such as rock music (Auslander 1999).

2.2.9 Virtuosity

> The border between virtuosity and gimmickry is a matter of convention and taste. (Siegel 2009, p. 200)

A *virtuoso* is an exceptionally skilled performer (Dobrian and Koppelman 2006). Kennedy (1996) writes that the description may carry an implication that a virtuoso performance excludes emotional or expressive artistry but this interpretation is rare within NIME research. Creating an instrument that will inspire a musician to dedicate the time and effort required to become a virtuoso is often an implied goal within NIME research (e.g. Wessel and Wright 2002; Jordà 2004; Hepting et al. 2005; Arfib et al. 2003).

Dobrian and Koppelman (2006) argues that virtuosity facilitates (classical) musical expression by freeing the performer from thinking about the basic functionality of the interface, allowing them to concentrate on listening and expression. They argue that the lack of virtuosity (i.e. instrument-specific technical skill) is the "elephant in the corner" of NIME research. Others have focused on the difficulty of *perceiving* virtuosity. Paradiso and O'Modhrain (2003) discuss the 'non-instrumental' gestures of many DMIs, challenging designers to create mappings that allow audiences to 'smell the digital sweat as they push their instruments to the edge.'

Fyans and Gurevich (2011) highlight the problem of perceiving a performer's skill without some kind of understanding of their intentions and what might constitute a mistake. We will return to this issue in Chapter 6. For now, we conclude that virtuosity is seen as a significant aspect of live performance and that there is a general dissatisfaction that DMI performances do not seem as virtuosic as performances with conventional instruments (Schloss and Jaffe 1993; Schloss 2003).

2.2.10 Intimacy, transparency and causality

Moore (1988) defines intimacy as *the match between the variety of musically desirable sounds produced and the psychophysiological capabilities of a practiced performer*. Paramount to this definition the subjective delay in the perceptual feedback loop between performer and controller. Moore considers the most intimate of controllers the human voice—a result both of the immediacy of our physical connection and the fact that our ability to control it is informed by our experience of speaking.

Fels et al. (2002) build on this definition to define transparency as the quality of a mapping indicative of *the psychophysiological distance, in the minds of the player and the audience, between the input and output of a device mapping*. Fels et al. argue that a stumbling block of DMIs is that, in part due to the possibilities that arise from the controller/instrument split, it is not immediately clear how the performer's actions relate to what is heard. The issue is confounded by a lack of the common cultural knowledge of their workings and gestures enjoyed by traditional instruments. Therefore, the onus falls upon designers to create transparent DMIs.

Note that transparency does not imply that the audience must understand how to play an instrument or even how it works. The term 'psychophysiological distance' is perhaps somewhat difficult to define and others prefer to consider the perception of a causal link between input and output (Schloss and Jaffe 1993; Schloss 2003; O'Modhrain 2011). Fyans et al. (2009b) describe the notion of transparency as simplistic, arguing that the spectator experience should be considered more in terms of perceiving intention, results and errors. We will discuss Fyans et al.'s model in more detail in Section 2.2.14.

Fels et al. (2002) argue that *metaphor* is a key tool in creating transparent mappings as it allows both player and audience to draw on familiar knowledge. Gadd and Fels (2002) present MetaMuse, an instrument consisting of a watering can that controlled a granular synthesiser through 'pouring' sound grains onto different parts of a palette. Other metaphors for control such as scrubbing and dipping are presented by Wessel and Wright (2002).

Creating mapping metaphors draws upon what Levitin et al. (2002) describes as our *cognitive maps* of how gesture relate to sound, a topic that was explored in Section 2.2.6. The key lesson to draw from this section is that it is a consideration for the *audience* as much as it is for the performer.

2.2.11 Controllability and diversity

Controllability is the degree of influence a (skilled) player has over the sound produced by their instrument. Wessel and Wright (2002) point out that control may be at a high level of abstraction, leaving algorithms to fill in trivial details, but the instrumentalist needs to *feel* that they are in control, much in the manner that a conductor might be an orchestra. Wanderley and Orio (2002) also consider controllability as a subjective measure but distinguish between feature controllability—what sounds can be made—and timing controllability—temporal precision over when these features are realised. Of course, *controllability* implies that the performer is able to produce these finer details in a predictable and reproducible manner. Thus we may distinguish *output complexity*—the range of possible outputs—from *controllability* or *diversity* (Jordà 2004).

Levitin et al. (2002) investigate this issue by looking at the beginning, middle and end of a musical event (generalising from the attack, sustain and decay of a note) and considering what means of control are typically afforded by different means of creating sound. Jordà (2004) considers the issue at a more abstract level, defining three levels of diversity.

Micro-diversity describes performance nuances, variations that could occur within a given piece without preventing it from being recognisable.

Mid-diversity describes how different two different pieces of music could be.

Macro-diversity considers the range of musical styles, contexts the instrument could play as well as its ability to assume different musical roles and accommodate a variety of musical partners.

Excluding, perhaps, a live coding context, these micro and mid levels of control need to be available without reprogramming the instrument to really be considered *diversity*.

Jordà's model is built around the classical model of expressivity, with micro- and mid-diversities respectively arising from the roles the performer and composer may take in a traditional concert. It is a powerful means of regarding the ways in which one interacting with a music system may see their role and potential opportunities. We may perhaps see it as a subdivision of the Sound category of Jo and Tanaka's matrix of musical participant (Section 2.1.9).

But in terms of ecological expressivity, it is something of a limited ambition. Does an instrument that embodies the musical ideals of its designer not in some sense limit the diversity of music an instrumentalist may create with it? Tanaka (2006) warns that thinking of instruments in such utilitarian terms risks turning them into 'tools', devoid of distinguishing characteristics or personality.

2.2.12 Learnability

Within the HCI literature, *learnability* is defined as the degree to which a system is easy to learn by the class of users for whom it is intended (Michelsen et al. 1980). Grossman et al. (2009) argues that there are two important types of learnability.

Initial learnability: Improvement within the first use of a system.

Extended learnability: Change in performance over time.

With conventional instruments, extended learnability tends to be of more interest, as it is accepted that a significant amount of time will be needed to develop proficiency (Lehmann 1997). However, for DMIs, potential players have less patience than they might for a normal instrument (Dobrian and Koppelman 2006; Weir 2012). It is generally accepted that instrumental mastery should require devotion and effort (e.g. Jordà 2004). But without an established instrumental practice demonstrating the worth of the instrument, those which grant a potential player *no* proficiency during their first experience (as might be conceivable at a first trumpet lesson) are unlikely to entice this devotion. This desire for easier early experiences with DMIs extends beyond pure necessity. Many desire to make instrumental music making more accessible as well (e.g. Blaine and Perkis 2000).

Our two types of learnability are equivalent to what Wessel and Wright describe as a *low entry fee with no ceiling on virtuosity*. Making things easy for novices often seems to limit the potential for more advanced users, leading to a 'toy-like character' (Wessel and Wright 2002). But Jordà (2002) argues that this trade-off is not inevitable (indeed, the Reactable (Section 2.2.2) might be considered evidence of this claim).

Jordà (2004) considers our two types of learnability more generally in terms of the learning curve. He considers the *efficiency* a player may have with an instrument as the controllable diversity of musical outputs mediated by the amount of effort required of the player to produce them. Thus, whilst the kazoo may offer a greater diversity than the piano in the short term, improvements in efficiency dry up whereas improvements in the piano's efficiency continue to develop (Figure 2.16). We will consider this claim with a more critical eye in Section 6.1 when we review the necessity of *perceiving* skill within a musical performance.

For many public space installations, potential participants may learn how something works by watching others interact with the work. It can be a social experience and designers may wish to consider how effective and engaging the vicarious learning process is (Reeves et al. 2005).

2.2.13 Explorability

Explorability is the degree to which a musical instrument facilitates exploration. Wanderley and Orio define it as a capability of a controller referring

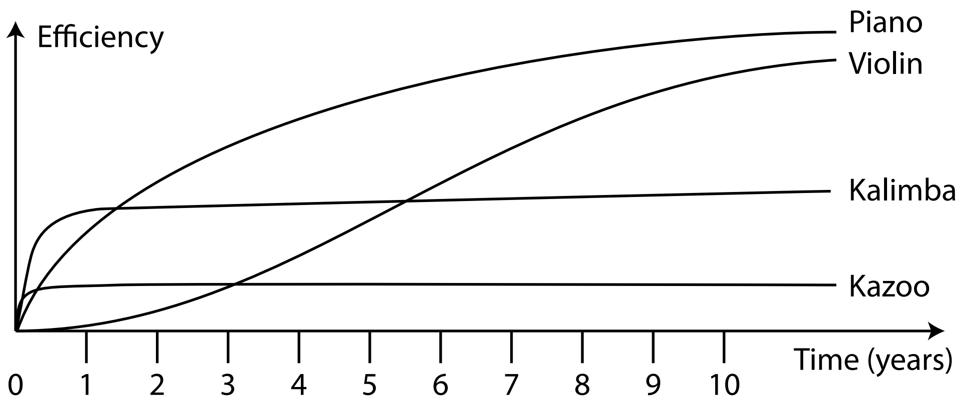


Figure 2.16: Jordà's learning curve estimates of the relative differences of the increase in relative efficiency over time between four musical instruments. The kazoo and kalimba (an instrument with a set of fixed pitch metal prongs that are plucked by the thumbs) offer easy short term gains. However, they quickly plateau in contrast to the piano or violin. Adapted from Jordà (2004).

to the number of different gestures and gestural nuances it may recognise (Wanderley and Orio 2002; Wanderley 2001b), what Jordà (2004) describes as *input complexity*.

However, this definition does not quite capture the essence of what we mean by explorability. As well as *permitting* a diversity of inputs, we want an instrument to lead us into experimenting and discover new aspects, to spark creativity (Gelineck and Serafin 2010). As Machover describes:

> While many have also been successful in designing controllers and interactions that "hook" a novice user, even in distracting, high-powered public spaces, few have been able to make such systems "nourishing" as well, capable of encouraging deeper exploration and continued discovery and creativity. (Machover 2002)

Gelineck and Serafin (2010) argue that for an environment to encourage exploratory behaviour it must be rich complex and somewhat mysterious but remain intuitive in order to give the user confidence to continue. Paine

(2002) makes a similar point. A participant develops a cognitive map of how it responds. The designer needs to balance confirmation of this map with scope for discovering new outcomes.

These observations are analogous to the learnability/controllability trade-off. However, instead of merely seeing the early stages of learning as a burden that must be overcome, it is seen as a part of the experience in and of itself. This idea underlies much of the theory that will be developed in Chapter 4.

Learning and exploration are recurring themes throughout this thesis. In Section 4.1, they will be revisited in a review of some prior work regarding exploration, play and curiosity.

2.2.14 The view from the audience

Although it has long been acknowledged that it is important to see a performer putting in a lot of physical effort (e.g. Waisvisz 1985), one area of NIME research that is often overlooked is the experience of the audience (Fyans et al. 2009b). (Exceptions include Fels et al. (2002), discussed in Section 2.2.10, and Schloss and Jaffe (1993).)

It is often cited that being able to perceive *causality* between the visual and auditory aspects of a performance is important (Schloss 2003; Paradiso and O'Modhrain 2003). Magnusson and Hurtado (2008) argue that audience understanding of a performer's activities is a fundamental aspect of musical performance as performer and audience need to share the same 'ergonomic space'. However, others argue for the need to perceive skill in a performance (Fyans and Gurevich 2011; Schloss and Jaffe 1993).

How interactive music systems are perceived by audiences and participants is a core theme of this thesis, and we will consider causality and the reasons it is necessary in more detail in Chapters 3 and 6. Here, however, we will review some contemporaneous research from Fyans et al. (2009a).

Fyans et al. (2009b) argue that there are five questions a DMI designer needs to consider for an audience to be able to perceive skill in its performance.

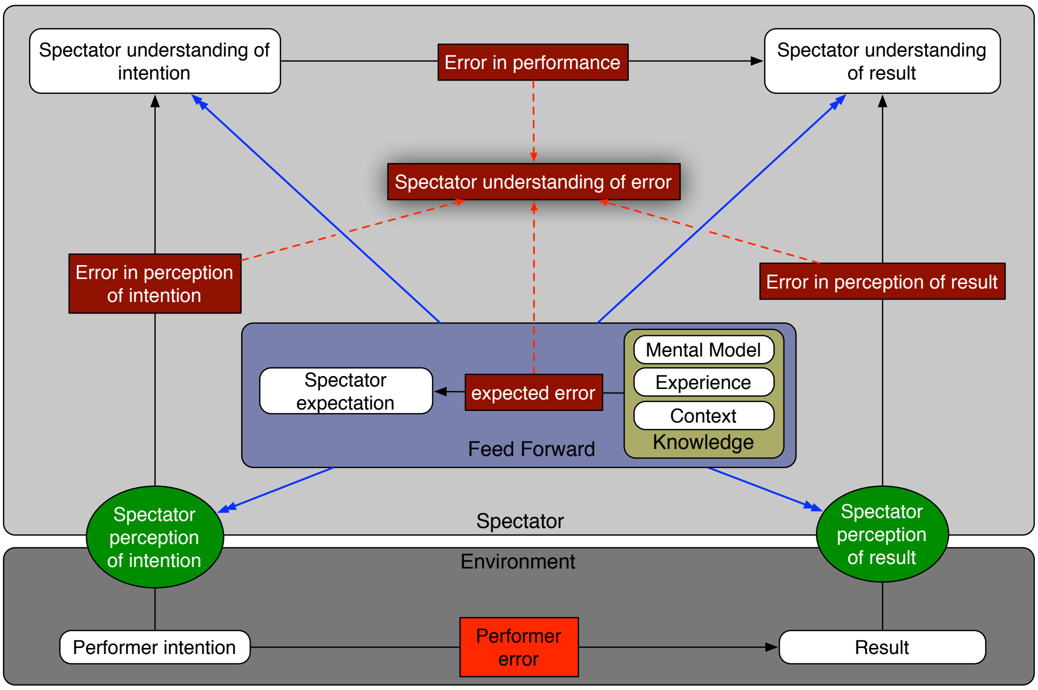


Figure 2.17: Fyans et al.'s model of spectator understanding of error. Error is perceived by a spectator as a combination of context, experience, their mental model of the instrument and their expectations. Reproduced with permission from Fyans et al. (2009b).

1. *Address: How does the spectator know that the performer is directing communication to the system?*

2. *Attention: How does the spectator know that the system is responding to the performer?*

3. *Action: How does the spectator think the user controls the system?*

4. *Alignment: How does the spectator know that the system is doing the right thing?*

5. *Accident: How does the spectator know when the performer (or the system) has made a mistake?*

For example, Fyans et al. (2009b) model of error (question 5) is shown in Figure 2.17 as a combination of the spectator's perception of the performer's intention, their perception of what was actually performed and their expectations based on their understanding of the instrument, other experiences and the context of the performance.

In a series of qualitative studies, Fyans et al. (2010; 2011) compared audience responses to four instruments: the violin, the sheng (a Chinese blown free-reed instrument), the theremin and the 'Tilt-Synth'. The Tilt-Synth was a DMI created specifically for the experiment with audience communication of intentionality in mind. Large obvious physical gestures were combined with fine-grained control through buttons, (mode-changing) switches, sliders and accelerometers. Both the sheng and the Tilt-Synth were unfamiliar instruments for all but one of the participants in question. Likewise, both instruments were performed by musicians unfamiliar with them. Whilst the sheng was perceived as requiring skill to use, the Tilt-Synth performance was considered to be lacking in effort: 'simple to control' and 'just pressing buttons' were some comments from participants. Fyans and Gurevich (2011) relate this discrepancy to the disembodied interaction with the Tilt-Synth. Participants could relate the sheng to their own physical experiences to gauge how much skill would be required to use it.

In Section 6.1.1, we will reconsider the audience's perspective and why many DMIs seem to struggle to gain acceptance in the concert arena. This will allow us to develop a new framework of Perceived Agency that considers in more detail interactions between performance context and a spectator's understanding of an instrument as well as different types of skill and the relevance of liveness.

We will be drawing extensively on the above topics when we consider the interactive music experience throughout this PhD, as well as in Chapters 3 and 6 where we will consider the audience experience of a NIME performance in more detail. However, as we have seen, much NIME research has been undertaken with a focus on the perspective of a performer with a concern for the sounds one might produce rather than the experience of producing them.

2.3 Enjoyment

In this section, we will briefly consider some broader research into why some activities seem more enjoyable than others. We will draw upon this material primarily in Chapter 4 when seek to understand how aspects of an interactive music systems can affect how enjoyable it is to explore.

2.3.1 Flow

Flow is a psychological state originally introduced by Csikszentmihalyi (1975, 2002), who described it as 'optimal experience'. The theory arose out of Csikszentmihalyi's findings that moments in which individuals reported themselves as happiest usually involved activities that were both challenging and required specific expertise, such as sport, playing music or creating art. Csikszentmihalyi and LeFevre (1989) collected further evidence correlating this balance of challenge and skill with more traditional measures of wellbeing in both the workplace and leisure activities.

The basic theory of flow is that a person's skills must match the challenge of the activity that they are pursuing (Csikszentmihalyi 2002). As an activity develops, its level of challenge should develop in line with the skills of the individual, creating a 'flow channel' of steady improvement in their ability (Figure 2.18).

Further research identified the need for both skills and challenges to be above average for the individual, else the result may be apathy rather than flow (Nakamura and Csikszentmihalyi 2002). Subsequent developments led to more complex models including boredom, relaxation, worry, anxiety, control and arousal as well (Figure 2.19).

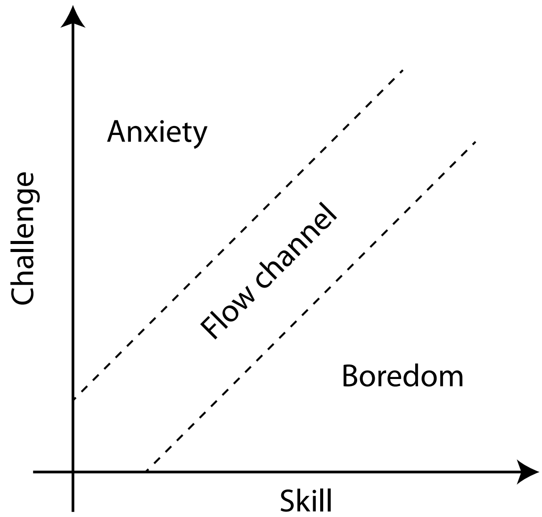


Figure 2.18: The basic flow channel defines a pathway of progression where the level of challenge and the level of skill remain matched. If they are mismatched, anxiety or boredom can result. Adapted from Csikszentmihalyi (2002).

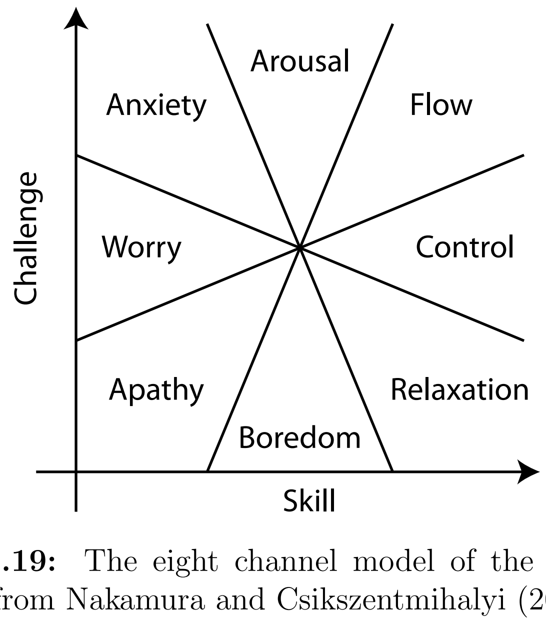


Figure 2.19: The eight channel model of the flow state. Adapted from Nakamura and Csikszentmihalyi (2002).

Flow is often described in terms of prerequisites that an activity must meet to elicit it (the *antecedents*) and the subjective *characteristics* of the experience when an individual is in a flow state (Pearce et al. 2005). Rettie (2001) describes the antecedents as

- Clear goals,
- Immediate feedback,
- Perceived skills/challenge matched and relatively high

and the characteristics as

- Merging of action and awareness
- Focused concentration
- Sense of potential control
- Loss of self-consciousness
- Time distortion (time passes more quickly)
- Autotelic experience (the activity is felt to be worth doing for its own sake)

The difference between an antecedent and a characteristic is not always clear. For example, Murphy et al. (2011) includes concentration as an antecedent alongside the condition that the individual is free from interruption. We may also consider whether an individual's opinions prior to an activity might influence whether or not it seems autotelic. Furthermore, the need for clear goals is at odds with reports of flow in more exploratory activities such as internet browsing (Rettie 2001). Ghani and Deshpande (1994) investigated flow in those using computers at work and found a *sense of being in control* to be more important than perceived skill, with challenge becoming more important to those with greater autonomy and feedback. They also found a significant link between flow and exploratory behaviour.

It is important to note that challenge and skill are *perceived* qualities. The individual must *believe* the task is challenging and that their skills are approparately advanced to deal with this challenge.

As we can see, flow can be a rather elusive concept. Research into the flow experience remains a lively field. Although it will not be forming a part of the core of our argument, there are three key insights to observe.

Firstly, there is a particular enjoyment in successfully performing activities that require skill. This notion will arise throughout this thesis, both in the Emerging Structures model developed in Chapter 4 and the framework of Perceived Agency introduced in Chapter 6. Flow has been identified as a contributing factor to motivating children to practise and improve their musical abilities (O'Neill 1999). Likewise, this argument has been applied to explain the attraction of increasingly more 'challenging' media within our favourite genres (Sherry 2004).

Second is the fundamental importance of *balance* underlying flow. This crops up when we consider how easy instruments are to learn in Section 2.2.12 and in more information theoretic terms when we overview expectation-based models of musical perception in Section 2.4.3.

Finally, we observe that it is *perceived* skill that must be balanced with *perceived* challenge. This suggests that our beliefs about the extent to which others may perform with similar aptitude may play a role in user experience. This idea will be expanded upon in Chapter 6.

2.3.2 Engagement and immersion

Other theories of experience consider *engagement* and *immersion*. Douglas and Hargadon (2000) distinguish between these two terms in the context of interactive narrative. Immersion is the pleasure of losing ourselves within the world it constructs, forgetting about the *frame*, i.e. the mechanics of the system that allow the user to interact with this world. Engagement, on the other hand, is the pleasure of relating the work as a whole to our understanding of the genre. Sherry (2004) describes this understanding of a genre as our knowledge of its *formal characteristics*. Works that fit neatly into out understanding of a genre may still give us the 'easy' pleasure of immersion but it is those that challenge us to make sense of them through violating convention—or our internalised understanding of convention—that may give us the pleasure of engagement. Although engagement and immersion may interfere with each other, Douglas and Hargadon propose that when occurring simultaneously, we are being challenged (engaged) but also achieving our goal (immersion). Thus, we have high challenge and high skill leaving us with the experience of flow (Douglas and Hargadon 2000).

Brown and Cairns (2004) coded interviews with seven videogame users following a grounded theory approach to investigate immersion. They developed categories of three progressive stages that a player descends through when playing a game.

Engagement is the initial stage where time and effort are willingly invested. The player is interested but not yet emotionally involved.

Engrossment happens when the player becomes emotionally involved in the game world. The player is less aware of their surroundings and stopping play can leave the player feeling emotionally drained.

Total immersion is the sense of being totally detached from reality to the extent that the game is all that seems to matter. The player feels present in the game and empathy with the character or team they are controlling within the game.

In a similar manner, Jennett et al. (2008) distinguish between *cognitive absorption*, *flow* and *immersion*. They describe immersion as an experience specific to videogames and find that increased levels of immersion increases as the time taken for the player to be able to reengage with the non-game world, measured through task performance.

As with flow, there are not universally accepted definitions of engagement or immersion. But there are interesting lessons to learn from the above research. The language used to describe immersion is similar to that of Fels's model of embodiment (Section 2.1.6) and in discussions of transparency (Section 2.2.10).

To extrapolate research focused entirely on videogames into interactive music requires caution. However, we should highlight the distinction observed between two key types of pleasure, that from *developing an understanding of a system* and that of *acting within the world defined by a system*. These are analogues to different types of exploration, which we will meet in Chapter 4.

2.3.3 Mutual engagement

Bryan-Kinns et al. describe *mutual engagement* as the points within a creative collaboration where *people spark together* and *lose themselves in their joint action* (Bryan-Kinns et al. 2006, 2007). It is similar to Sawyer's (2006) notion of *group flow* except more specifically, there is engagement from a participant with both the product of the activity as well as other participants (Bryan-Kinns and Hamilton 2009).

Although the work of this thesis has been developed with a focus on individual experience, the topic of mutual engagement is relevant when we create and evaluate a collaborative IMS in Chapters 7 and 8. We will explore some of the more social issues surrounding musical creation in Chapter 6.

2.4 Musical perception

In this section, we will briefly outline some current research into how and why we appreciate music. Although a detailed overview of musical perception is beyond the scope of this thesis, the results outlined this section are fundamental to the development our model of exploration in Chapter 4.

In particular, we focus on two concepts that seek to explain how we interpret music, why it engages us and how it is able to move us emotionally.

2.4.1 Communication

Livingstone and Thompson (2009) argue that music plays an adaptive role within human evolution through allowing emotions to be communicated and consequently mutual understanding and instruction. As a result, within music we may hear mimicry of innate human noises such as laughter, crying and screaming (Moorer 1972). Small (1998) argues that the ritual, sounds and gestures of music provide for a type of non-sequential implicit communication that language does not. This serves the purpose of allowing groups of individuals to establish and communicate their relationships and experiment in different roles.

Musical Acts

Murray-Rust and Smaill (2011) create a logical model of Musical Acts to represent how a group of musicians establish a common context through musical actions. They model a musical performance as the progression through a directed lattice of states. The position of a state within the lattice is determined by which states it subsumes and which it is subsumed by, with subsummation determined by the set of constraints imposed by a state. For example, a state which determines only a root note of C will subsume a state that determines a chord of C minor (C E♭ G), which itself will subsume a state that determines the chord of C minor 7 (C E♭ G B♭). At a given point in time, different agents may be in distinct states and the dynamics of a group of musical agents are described as a negotiated journey through this state space. A musical action is then described as the intentional and intelligible change in state by an agent. Each action of a given agent A playing with another agent B is given a *signature* characterised by three relationships:

how A's new state relates to B's state, how A's old state related to B's state and how A's new state relates to A's old state.

Murray-Rust and Smaill provide a convincing model of how the communication and exploration of roles described by Small might take place. Considering a musical act as a process of identifying constraints over the current and potential future states through interaction is similar in nature to the model of exploration I will introduce in Chapter 4. Whilst the context in which their model is formulated is tied closely to jazz improvisation, it relies on an externally defined set of musical 'facets' (e.g. chord, dynamics, tempo) and corresponding functions within an agent to interpret values of these facets from what is heard. Doing so abstracts away many subjective details leaving a general model of the underlying musical processes. However, in Chapter 6 we will see that across musical practices there are other processes by which intention is communicated.

The Musical Acts model focuses entirely on how sounds are produced and interpreted within an existing music theory. It assumes that this theory is to a large extent available as common knowledge to the players. However, in Chapters 3 and 4, I will argue that interactive music musical sounds should be considered concurrently with musical gestures. We will be using the term *musical actions* to mean, not just how creating sound may influence a shared performance context, but in the more literal sense of describing how a gesture (i.e. an action) creates sound. Furthermore, our scope of enquiry is focused on non-expert first-time users. Therefore the model of Emerging Structures to be presented in Chapter 4 may be considered as describing the more fundamental exploratory act of how an individual may arrive at such a theory in the first place.

Musical Acts is based on Speech Act Theory (SAT) (Searle 1969; Herzig and Longin 2002), which reasons about how words are spoken with the intention of changing the state of the world rather than to communicate a description that may or may not be true. For example, the statement 'the door is always open' interpreted as a *direct utterance* communicates an assertion of a fact. However, the statement interpreted as an *indirect utterance* may be an act granting an individual permission to enter a building. In particular, Murray-Rust and Smaill draw upon Herzig and Longin's (2002) logic of intention, which is based on the belief-desire-intention framework: an agent, based on their belief about the state of the world and their desired final state of the world, forms intentions—actions that the agent has planned and committed to in order to reach a subgoal state that helps attain their desired state of the world.

Interestingly, in Murray-Rust and Smaill's computational implementation of their model the agents choose actions based on not a desired final state but the signature of action they would like to perform (i.e. whether their action subsumes, is subsumed by, alters or deviates from other players' states). Once a signature is chosen then musical values are sought to create an action with this signature. We might therefore argue that the Musical Acts model as implemented by Murray-Rust and Smaill is more experience-focused than SAT.

Speech Act Theory will be revisited in Section 9.1.1 as a comparative example to the theoretical material presented in this thesis.

2.4.2 Representation

There is also evidence of the importance of the *representational* aspects of sound: that the auditory building blocks with which we perceive music have their origins in our ability to perceive and understand our physical environment (Huron 2006). Hearing involves unconsciously identifying the physical origins of sound (Van Nort 2009). However, the *mimetic hypothesis* (Cox 2001) further proposes that when an individual hears a sound they imagine how it is created. This may potentially happen unconsciously through the mirror neuron system (Brent 2011). Even for non-causal sounds (those without a physical origin), there is evidence that listeners associate them fairly consistently with types of gesture (Caramiaux et al. 2011). Indeed, in acousmatic music where the original origin of a sound remains hidden, Windsor

(2011) argues that it is the puzzle arising through this inability to identify a cause that engages a listener.

2.4.3 Expectation

Expectation has been long established as playing a fundamental role in musical perception (Meyer 1961). Although a detailed analysis is not necessary, the models we outline here lay crucial groundwork for the Emerging Structures model of exploration presented in Chapter 4.

The Implication-Realization (I-R) model

Narmour (1990, 1992) proposed with his *Implication-Realization* (I-R) model of melodic expectancy that such (unconscious) expectations may arise both innately and due to cultural exposure (Schellenberg 1997). Narmour's complex model originates from the two simple axioms

1. Repetition establishes expectancy of a further repetition,
2. Novelty establishes expectancy of further novelty.

These future expectations are described as *implications* (Narmour 1992). Where implications are met (*realisation*) or denied, a sequences of notes may be *closed* and subsequently form a part of a larger structural grouping (Cross 1995).

Narmour considers innate implications over a range of melodic features including interval and register and establishes a number of *principles*. Implications may arise from innate and universal principles, from within a work (*intra-opus*) or from stylistic knowledge (*extra-opus*) (Pearce and Wiggins 2004). Through applying the principles concurrently to different features and at different scales—in combination with intra- and extra-opus implications—potential conflicts and complex implications may arise. Lerdahl and Krumhansl (2007) consider expectation, and thus the anticipation of its realisation, as one of three components of musical tension alongside dissonance and harmonic instability relative to a tonic.

Huron's 'ITPRA' model

Later statistical and empirical evidence has suggested that the complex innate principles of the I-R model may themselves be explained as learnt rules through musical exposure (Pearce and Wiggins 2004). Huron (2006) proposes that music may elicit emotion through the interactions of different aspects of the brain's need to predict its environment. Huron's model is called *ITPRA*, named after the five stages of expectation it describes. Before the event we have:

1 Imagination: Our continuous expectations regarding our environment.

2 Tension: We may increase our attention and arousal in line with the expected outcome. Tension describes the period spent in this state prior to the event.

And then following the event:

3 Prediction response: An unconscious limbic reward or penalty determined by the accuracy of the prediction

4 Reaction response: Neurologically reflexive response that assumes the worst case outcome.

5 Appraisal response: A more complex and potentially consciously informed appraisal of the outcome that determines any negative/positive reinforcements.

The model rests on the assumption that at a low level within our perceptual system, predictability and surprise result in positively and negatively valenced emotion respectively. However, if the reaction or appraisal of a (negative) surprise is subsequently positive, then there is a *contrasting valence* which heightens the positive. Under this model, when a composer builds up to a climax, they are establishing ever-growing anticipation of a negative outcome which gives the potential to amplify a subsequently positive outcome.

Expectations in the model are derived through simple statistical models and lifelong musical exposure. Huron's builds on Narmour's (1990; 1992) intra- and extra-opus sources of expectation to consider four types.

Schematic: Relating to commonplace musical expectations and stylistic norms (similar to the *formal characteristics* of Section 2.3.2).

Veridical: Prior knowledge of the specific work itself.

Dynamic: Expectations formed from hearing the piece until this moment.

Conscious: Consciously established expectations.

Huron's model has been criticised for muddying the difference between conscious and unconscious expectations: although conscious expectations play little role in his model, many of the conclusions of the model draw on evidence regarding the conscious expectations (Benjamin 2007). However, a strength of the model is its potential to combine the expectational with the representational in the prediction and reaction.

Information dynamics

Information dynamics uses the tools of information theory to describe the evolving expectations present in music. Structure in music is considered a *redundancy* in the signal—a pattern that may be used to reduce uncertainty of the future signal. Representing a musical signal as a sequence of events, listening is considered in terms of how each event affects the listener's understanding of this structure (Dubnov et al. 2011). Therefore, rather than simply considering entropy (i.e. the information content), information is considered with respect to a listener's existing expectations of what they hear.

Dubnov (2006, 2010) models this in terms of the 'information rate' which he defines the mutual information between past events and a current event—how informative what we are hearing is *given* the predictions we could have made from past knowledge.

However, for Abdallah and Plumbley, the salient question is not 'how far has the current event aided our understanding of the present state of affairs?' but 'how far has the current event improved our ability to predict future events?' This property they describe as the *instantaneous predictive information rate*. Averaging over the signal then provides the *average predictive information rate* (APIR). We may then combine expectation of the next event with previous experience of predictive information to produce the *expected predictive information*: how far we expect the next event to improve our ability to predict future events.

The APIR moves beyond simple measurements of predictability to consider the rate at which redundancy within a signal is becoming perceivable and thus the rate at which our ability to predict the signal is improving. It is minimised by both random noise (i.e. a signal that will *never* be predictable) and known repetition (i.e. a signal that is *already* maximally predictable). In this sense, the APIR is a measurement of the rate at which the structure of apiece is unfolding.

If we consider a random signal to be 'too challenging' to predict and a predictable one to be 'too easy', then the APIR allows us to connect flow (see Section 2.3.1) with musical perception. Just as flow implies that enjoyable activities are those in which challenge is optimised for the development of skill, information dynamics implies that enjoyable music is that in which unpredictability is optimised for the development of predictive ability.

These balances are both instances of the Wundt curve (Berlyne 1970), an inverted U-shaped relationship between complexity and what is described by Berlyne describes as *hedonic value* (Figure 2.20).

We have seen the perceptual enjoyment derived from listening to music as a process of establishing implications to be either realised or subsumed into a yet more complex implication. Huron (2006) sees the listening process as a continual bolstering of sets of exposure-based models which then vie to

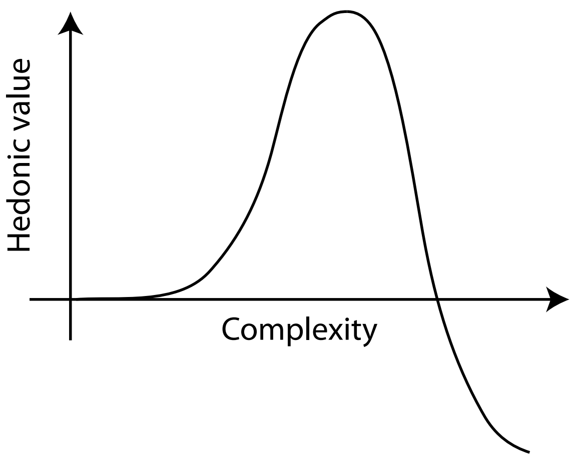


Figure 2.20: The Wundt Curve relates the complexity of a stimulus with its hedonic value. Adapted from Berlyne (1970).

anticipate the next event. The picture painted by Narmour's (1990; 1992) I-R model is somewhat less dependent on temporal progression: implications for the future are presented as missing pieces of a greater structural whole. With Abdallah and Plumbley's (2009) model we see something of a parallel to Csikszentmihalyi's (2002) theory of flow: the ability to improve is maximised by balancing the complexity of the task with the *current* ability of the individual.

These parallels are in different domains: the APIR describes our ability to *understand* our environment whilst flow considers our ability to *act* upon it. In Chapter 4, we will consider how these two domains in terms of exploration where we are both acting in order to understand *and* understanding in order to act.

Uniting our expectational models is the notion of piecing together a structural whole. It is similar to Minsky's (1981) theory that hearing music exercises the same metal capacities as viewing scenery: the eye darts around in order to establish a consistent model that may accurately predict what is seen. In music, as in cinema, our access to the full picture is restricted by the composer (or director). As we consider the perception of interactive music in Section 4.2, this analogy of exploring a scene to create a model will become more prominent.

It has been important to note that underlying our expectation-oriented models are communicative and representational acts (Sections 2.4.1 and 2.4.2). In Chapter 3, we will begin to consider music not just as sound but as *gestures creating sound*. Small (1998) argues that whilst much of sounds within music are somewhat arbitrary and abstract, the language of gesture—both as a means to communicate abstract ideas and to represent the physical world—remains more representational and culturally invariant. We will see this most prominently in Section 3.3.1 when we consider the composed instrument as a joint composition of both gesture and sound.

2.5 Evaluation methods

In this section we will outline the different evaluation methods commonly used in this and related fields. This will provide the context against which to compare our own evaluations in Chapters 5 and 8.

Any kind of evaluation of a creative tool will need to determine the context of use and whose response is of interest. Examples of subjects of interest include: an audience of a performance, a performer, a participant of an installation, the creator of a work, a customer or a judge with domain-specific expertise (O'Modhrain 2011).

Conventional research into Human-Computer Interaction (HCI) has been concerned with use-case scenarios that focus on performing tasks. This *task-oriented design* allows an interface to be evaluated by the efficiency with which tasks may be performed (e.g. Carroll and Carrithers 1984).

Software may also be evaluated by the experience of the user whilst using it rather than the outcome of the tasks they perform. This focus is traditionally more familiar to game designers (e.g. Sykes 2006). However, as recreational use of computers has increased it become a more prominent topic within HCI (Wright and McCarthy 2010). This second focus is referred to as *experience-centred* or *experiential* design.

2.5.1 Formal and informal evaluation

The formality of an evaluation is the extent to which it is conducted and presented in a rigorous fashion with a structured route from data to results (Stowell 2010). Formalising an evaluation procedure allows data to be collected and presented in a manner that ensures the conclusions drawn are supported by evidence beyond a researcher's personal opinions. As such, they are important in understanding the extent to which results may be generalised (Stowell et al. 2009).

Generally, most evaluation within NIME research is informal (Marquez-Borbon et al. 2011). This may be because the aims of the research are to produce a particular artistic result and therefore there is little objective criteria against which it may be assessed (what Nic Collins calls 'playing the composer card' (Eigenfeldt 2009)). It may also be due to the time commitment required to conduct formal evaluations. For example, Stowell (2010) reports that his qualitative analysis took approximately 27–36 hours to analyse per hour of interview in addition to the time required for transcription. However, informal evaluation increases the potential risk of results that have been 'cherry picked' to support the researcher's personal agenda, whether intentionally or not. Without an understanding of how the researcher may have been led to their conclusions by the data, we cannot verify their reliability. This risk increases when the evaluator is also serving as creator.

Nonetheless, informal evaluation has a useful role to play (Stowell et al. 2008). In particular, informal analysis of qualitative data may be used to help explain the significance or insignificance of a quantitative result. In this thesis we conduct a formal quantitative analysis in Chapter 5 and use informal qualitative analysis to reason about which direction to take subsequently. We then perform a second formal qualitative analysis in Chapter 8.

2.5.2 Subjective and objective evaluation

Any method where people are used instead of a tool as means of measurement is a *subjective evaluation* (Dix 2004, p. 358). This is common within our field as we are often interested in understanding experiential or aesthetic qualities. Subjective evaluations may be criticised on grounds of *reliability* as participants' responses may easily be influenced by factors other than the one we are testing. To overcome this, we may use an objective measurement as a more reliable indicator, such as measuring the number of people who visit an exhibition or a participant's galvanic skin response. However, one may then call into question whether what has been measured is actually indicative of the quality we are trying to measure, giving grounds to criticise the *validity* of the results.

2.5.3 Quantitative subjective methods

Quantitative subjective evaluation typically means asking people to fill out questionnaires to provide a comparative or quantitative answer to a set of questions about an experience the person has just been involved in (e.g. Rutherford and Wiggins 2002; Fencott and Bryan-Kinns 2010). However, it may also ask them to provide a continuous response to a stimulus through an input device (e.g. Melo and Wiggins 2003). It may be done in the lab (i.e. a controlled environment) (e.g. Rutherford and Wiggins 2002; Melo and Wiggins 2003) or in the field (i.e. our determined usage context) (e.g. Morrison et al. 2007). They are typically asked of a potential participant (e.g. Fencott and Bryan-Kinns 2010), performer (e.g. Hsu and Sosnick 2009), audience member (e.g. Rutherford and Wiggins 2002) or an expert judge (e.g. Wiggins et al. 2009).

Some research uses tests inspired by the Turing Test (Turing 1950), which tests a machine's intelligence by the extent to which it may convince a human that it is also a human (e.g. Ariza 2009; Robertson 2009; Murray-Browne and Fox 2009). This is arguably also a subjective method.

2.5.4 Quantitative objective methods

A common objective method used is to ask the user of an interface to perform a task, for example using a musical interface to recreate a sound, and then score their performance using a predetermined metric (e.g. Hunt et al. 2003; Wanderley and Orio 2002; Kiefer et al. 2008). Other approaches include measuring biometrics such as a subject's heart rate or galvanic skin response (an indicator of stress) (Hazlett 2008) and analysing system logs to measure differences in behaviour (e.g. Fencott and Bryan-Kinns 2010; Bryan-Kinns et al. 2007).

2.5.5 Qualitative methods

Qualitative methods are always subjective as they rely on the researcher to decide how to interpret the data (Denzin and Lincoln 2005). In our field, qualitative data is frequently published without any kind of formal analysis, often as an aid to interpret quantitative data that has undergone more rigorous analysis. However, rigorous qualitative methods of collecting and analysing data exist, some of which are outlined below.

The specific method used in any qualitative approach is often quite open. There is a diversity of methodologies used, ranging from informal opinions summarising an interview to formalised processes of extracting meaning from text and verifying their reliability. The onus remains with the researcher to convince an interested reader of the quality of their analysis by providing detailed documentation of the procedures they have followed (Yin 1994).

Discourse-based approaches

Qualitative methods often involve collecting written or spoken data directly from the party of interest. This may be collected after a specific experience through open-ended questionnaires and interviews or through *think aloud* methods where the participant is asked to provide a running commentary as they use a system (Hoonhout 2008). *Video-cued recall* is a further technique where the participant provides a running commentary of a video recording of their use (e.g. Costello et al. 2005; Healey et al. 2005).

The resulting *discourse* may then be *coded* where the material is organised into chunks or segments of text followed by the identification of themes or descriptions (typically around five to seven) (Rossman and Rallis 1998; Creswell 2008). These provide a foundation from which to form an interpretation.

Discourse analysis (DA) involves a very close reading of the text to identify specific features of the text such as repertoires (e.g. Inskip 2010), conceptual maps (e.g. Stowell et al. 2009) or rights and obligations (Banister et al. 1994, ch. 6). DA methods will be covered in more detail in Chapter 8.

Other formal approaches

Observation is the process of watching participants interact with a system and may be done live or using a video recording. As with discourse, observation is often accompanied by extensive coding procedures in order to make sense of what has been observed (e.g. Costello et al. 2005; Healey et al. 2005).

Ethnography is an approach to data collection where a researcher spends a considerable amount of time alongside a group of individuals recording their observations. It is usually conducted in the field and studies the culture of that group as revealed through their actions (Preissle and Grant 2004). It is particularly useful in understanding how a particular system is used by a set of users (e.g. Barthet and Dixon 2011) or how an issue such as technology has an impact within a specific context (e.g. Stowell and Dixon 2011).

'Autoethnographic' approaches involve applying ethnographic practices to the researcher. Examples include Magnusson (2011) who documented his experience of learning of a new live coding language (see Section 2.2.4), and Sudnow (2001) who wrote of his experiences learning to play jazz piano.

Grounded theory is an approach where theory is derived from data through systematic coding. The approach requires the researcher to suspend their prior knowledge and 'let the data do the talking'. A more detailed description is provided by Inskip (2010).

Evaluating the aesthetic and creative experiences

Both creative and aesthetic experiences are inherently subjective (Höök et al. 2003). Not only does mood affect how positively we regard a work, but mood affects right from the outset the cognitive processes we adopt when perceiving art (Forgas 1995). Leder et al. (2004) has argued as a result that research into experimental aesthetics should measure the mood of the participant before any study and consider it as a factor. Such an approach may be suitable for investigating aspects of the cognitive process within the psychology of aesthetics. However, when the goal is to understand the greater impact of design decisions within interactive systems, we may well argue that it is not enough to control for individual differences.

A similar argument applies when evaluating the potential of a work to facilitate a creative experience. Stowell and McLean (2011) argue that artistic creation is a rich and open task. As a result, reduction into individual measurable components for the sake of evaluation is likely to disrupt the very essence of the activity. Furthermore, our dependent variable is subjective. As such, we cannot measure it without risking changing it in some way. This is especially a problem in the cognitively rich tasks that our domain involves as interrupting actions interrupts thoughts which may disrupt the experience (O'Modhrain 2011). Stowell et al. (2009) point out that there is a particular danger in attempting to use think-aloud methods (see Section 2.5.5) during musically creative tasks as similar parts of the brain are active in both language and music processing. Indeed, it is not unreasonable to assume that even asking somebody to *think* about a particular aspect of their experience will alter it in some way.

We may resort to objective biometric indicators such as galvanic skin response or piloerection (causing goosebumps) (see Section 2.5.4). However, whilst potentially reliable indicators of specific experiences, it is limiting to reduce the complexities of the creative and aesthetic experiences to such simplistic measurements. After all, the 'chills' associated with piloerection may correlate with an emotionally poignant musical experience, but they may also be induced by the sound of fingernails scratching a blackboard (Grewe et al. 2011).

Each of the above evaluation methods seeks to understand the interaction between participant and system in some way. Interactive systems designed with aesthetic or creative outcomes in mind present a particular challenge when it comes to evaluation. We will discuss the advantages and disadvantages of qualitative research methods in Chapter 8. However, it is worth mentioning here that as a considerable amount of subjective analysis is performed by the researcher, it is less suited as a means to *verify* a preconceived theory. Its strength lies as a tool to develop a deep understanding of a particular interaction and how it relates to an individual's thoughts and behaviour.

2.6 Conclusion

We have covered a diverse range of subjects, primarily from the fields of NIME and interactive art but also looking at further research into flow and similar experiences, and musical perception. What appears as a common theme throughout this material is that learning is enjoyable, whether it is learning how an instrument or interactive music system (IMS) works, what it does, what it lets *us* to do, what it can teach us about ourselves—or unconscious learning as in our expectation-oriented models of musical perception. Particularly enjoyable seems to be learning that improves our ability to act—either to express ourselves, to create or simply to master a system. It is this combined pleasure of learning and creating that we will be working with throughout this thesis. However, we have also seen that creating a system that invites a participant to engage, explore and motivate themselves to learn is challenging. Making things easy to learn seems to risk offsetting the pleasure of doing so. What is it that makes a system continually engaging and exciting to use—in other words, *captivating*?

A common goal that is sought to attempt to achieve this appears to be in moving beyond simply providing control of a system to more complex interactions such as that of a conversation. However, we saw that the objective descriptions of directions of information flow within systems struggled to really capture this complexity. Throughout this thesis, we will be tackling this issue from a subjective standpoint. By this we mean our concern will be not only what happens but how these happenings are perceived, how they then affect what the participant does and, ultimately, how this perspective may inform the way in which we design IMSs.

We have seen that IMSs lie at the intersection of NIME and interactive art. A key difference in approach that has already become apparent between these two disciplines is that NIME research has typically focused on making systems for performance, whereas for interactive art the users are themselves the audience. This leaves the participant of an IMS in something of a mixed role between performer and audience member. Ideally, we would like to combine the best of both of these roles into an IMS.

In the next chapter we outline a practice-led project that will allow us to examine what it is that makes NIME performances unique to spectate. This will lead us towards an understanding of how interactive music can be more than simply music that is interactive, how we might combine the act of creation with the experiencing of another's creation and thus how to create a captivating interactive music experience.

Chapter 3

Musical interaction as an expressive medium

In this chapter I demonstrate how musical interaction may be considered an expressive medium in and of itself. I do so through a consideration of composed instruments*, Digital Musical Instruments created as expressive artefacts in themselves rather than simply tools to facilitate musical expression. I explore this issue with the Serendiptichord, a wearable instrument for dancers I created in collaboration with the artist Di Mainstone. We reflect that as well as defining a mapping, a composed instrument is an arrangement of musical interactions into a performance. As such, consideration needs to be given as to how they are perceived by an audience. We examine two aspects of this in particular: what the audience knows at the beginning of the performance and the rate at which a their understanding subsequently develops. We conclude that an understanding of the role of these aspects within the participatory context will form an important part of our understanding of a captivating Interactive Music System.*

> In this case, there may be numerous ways to characterise the aim, or multiple complex aims, but nonetheless, the DMI is a

> mediating technology, not an end in itself. After all, *instrument* is nothing more than a glorified synonym for *tool*.
>
> (Linson 2011, p. 422)

> Traditional acoustic instruments are never confused with tools.
>
> (Tanaka 2006, p. 272)

In the previous chapter we saw that whilst a Digital Musical Instrument (DMI) performance involves a performer creating sound through a computer for the benefit of an audience, the Interactive Music System (IMS) combines these two roles into a single participant. We also saw that whilst the DMI paradigm dominates research into New Interfaces for Musical Interaction (NIME), the line between what is an instrument and what is a musical work can become blurred, with some using the term *composed instrument* to describe this hybrid state (Section 2.2.2). As an IMS may be seen as a musical work requiring participation (Section 1.2), we may therefore see it as being closer to a 'non-expert composed instrument' than a 'non-expert DMI'.

Drawing a distinction between the composed instrument and the DMI may seem somewhat abstract. However, in this chapter we will demonstrate that these two paradigms are in fact quite distinct approaches to creating performances with practical consequences for how they are designed, performed and perceived. Key to these will be the concept of creating a musical work out of musical interaction itself rather than purely out of music. In subsequent chapters, we will then consider how to apply these lessons in the context of interactive music.

Our approach in this chapter will be through reflective practice. In Section 3.2 I will describe an exploratory project in which I collaborated with an artist to create a musical instrument for dance performance. We will then use this as a vehicle to explore why a composed instrument performance differs from that of conventional music. In line with our subjectivist stance (see Section 1.3), the focus will be on how such an instrument is perceived by an audience, how this perception is affected by performance decisions. However, perhaps paradoxically, to understand what an audience is there to spectate we need to consider what a designer is trying to present when creating an instrument.

3.1 The motivations behind DMI research

Following our review in Section 2.2 of some of the wealth of NIME research, we might consider the desire to make DMIs as 'good' as conventional instruments as being a common motivation. From this perspective, creating DMIs is a conventional engineering problem: success is assessed through potentially reliable indicators, difficult to measure as they may be. As useful and valid as such approaches are, we will here argue that this approach is not necessarily the best means of evaluating all musical instruments.

To do so we will consider motivations driving DMI creation in the context of three categories as follows.

Utilitarian motivations: creating DMIs that may perform better within an established role.

Utilitarian/artistic motivations: creating DMIs that are functionally suitable to achieve a particular artistic aim

Artistic motivations: creating DMIs as an act of personal expression.

Note that these categories are not necessarily intended to be exhaustive or mutually exclusive. However as we explore below, much of the DMI research described in Section 2.2 resides primarily within one of the above. We shall consider them each in turn.

3.1.1 Utilitarian motivations

We describe as a utilitarian motivation a desire to make DMIs that are 'better' at allowing a general instrumentalist express themselves musically. DMIs created for utilitarian motivations are typically evaluated through consideration of conventional musical instruments, identifying features that are desirable for all instruments (Wessel and Wright 2002; Jordà 2005) and identifying objective indicators of these (Wanderley and Orio 2002). Much of this is driven by a sense that DMIs have not been accepted into the concert arena by musicians or audiences; this is an issue we will consider this issue in more detail in Chapter 6. Here we consider what necessitates the need for DMIs in the first place?

Instrument usability

Expressing oneself through music is an immensely rewarding experience (Nachmanovitch 1990) but traditional instruments are difficult to learn to play, often requiring years of practice from an early age (Wessel and Wright 2002). Making the experience of creating music more accessible is often cited as a motivation for better DMIs (e.g. Hunt and Wanderley 2002), in particular the experience of collective music making (e.g. Blaine 2006; Jordà et al. 2005; Tanaka 2007). In fact, the implications of the term *non-musician* that only some people are sufficiently skilled to be creating music has been identified as a peculiarity of Western musical conventions (Small 1998). For some, the ability of DMIs to 'democratise' this experience is among their most exciting aspects (Chadabe 2004). Without the physical constraints of acoustic instruments, DMIs offer primarily two avenues to improve accessibility.

The first is the 'easy' musical instrument—such as the intelligent instruments of Section 2.2.4. The idea here is to let the computer take care of the difficult technical aspects and allow the performer to focus on expressing themselves. This may be beneficial in a therapeutic context, for example in special needs education (Challis 2011), or for home entertainment (Chadabe 1997). Such research may also be applicable in an installation context where a non-specialist audience might engage in a momentary interaction (e.g. Paradiso 1999). However, in the context of musical performance, the idea of removing the need for skill leaves many dissatisfied (O'Modhrain 2011). For example, we saw in Section 2.2.9) the argument that virtuosity is a necessary precursor of expressive musical performance. Others, however, have questioned this presumption (Jordà 2002).

More common, perhaps, is the desire to make DMIs more learnable but without limiting their potential for virtuoso use, which was discussed in Section 2.2.12. However, elsewhere learnability is presented not as *necessitating* new DMIs but simply as a necessity *of* new DMIs: we cannot expect a musician to devote as significant an amount of time learning an untested and unknown DMI as they might were they to take up the violin (Wanderley and Orio 2002).

As we are considering systems created for non-experts (Section 1.2), it will be necessary to consider the extent to which skill is a necessary component of expressive musical interaction. We shall do so in Section 6.1.

Instruments for computer music

Other arguments for creating DMIs focus on facilitating the performance of music that previously could only be composed on computer and rendered directly to audio. A key area where we see this is in 'laptop performance,' where software such as Ableton Live is used to add 'liveness' to computer rendered sound. There is a sense of dissatisfaction that much live computer music seems to be performed with the computer keyboard (Widmer et al. 2007) or generic MIDI keyboards (Jordà 2004).

As well as performance of prewritten music, others have sought to expand the possibilities of improvisation of computer music (e.g. Jordà 2005; Nicolls 2010; Dudas 2010). Improvisation provides those creating music an experience distinct from composition or rendition (i.e. performance of an existing work) (Nachmanovitch 1990).

Considered under the lens of concert expressivity (see Section 2.2.8), the difference between improvisation and rendition may seem superfluous from the audience's perspective. When we consider skill from a more ecological perspective in Chapter 6, we will explore why improvisation, rendition and a tape recording may lead to distinct perceptual experiences for an audience. However, at this stage in our argument it is only necessary to identify that there are established utilitarian needs motivating much DMI research.

As we saw this throughout our review of NIME research in Section 2.2, utilitarian motivations tend to lead to potentially measurable improvements criteria such as transparency, controllability, diversity, learnability.

3.1.2 Utilitarian/artistic motivations

We consider motivations within the utilitarian/artistic category as those that still seek to create DMIs as a means to facilitate the performance of expressive music but without a fixed idea of what this music or its performance context might involve.

Creating music to play on a new instrument and creating instruments to play new music are two sides of the same coin. Throughout the development of Western music, the development of new music has corresponded with that of new technology, and vice versa. For example, during the 19th century extra keys were added to the flute as chromaticism developed; today new means of continuous control are being developed alongside newly accessible timbral spaces (Kvifte 2011). As such, one may argue that evaluating DMIs by comparison with traditional instrumental role may unduly disregard other aesthetic possibilities (e.g. the 'glitch' aesthetic (Gurevich and Treviño 2007)).

Any musical instrument has a range of potential sound output. But inherent in any interface are suggested paths to follow, if not explicit then implied through ease of access (Rokeby 1998) and the metaphors embedded within the system (Fiebrink et al. 2010). Correspondingly, a new interface can trigger new ideas and new ways of thinking (Magnusson 2005; Jordà 2004).

As well as creating new types of music, DMI research may also be motivated by a desire to create new practices of music making. We saw this, for example, with Jo and Tanaka's (2009) matrix of musical participation (Section 2.1.9). Other examples include collaborative music making in shared virtual spaces (e.g. Fencott and Bryan-Kinns 2010; Bryan-Kinns 2012) or remote locations (e.g. Tanaka and Bongers 2001) and interactive dance (Section 2.2.4).

As instrument design becomes more about exploring new types of music and music-making, it starts to become an artistic as well as an engineering process. Whilst the instrument is still a tool that facilitates an artistic outcome, as the final outcome is ultimately dependent upon aesthetic preference there is no longer a uniform criteria by which to evaluate success. Imposing criteria derived from conventional performance practice is effectively limiting the scope of new instruments to existing practice. Ultimately, evaluation presents difficulties because we do not yet understand the potential artistic outcome to which the instrument may be of service.

3.1.3 Artistic motivations

Finally, we consider those who create DMIs not merely as a means to facilitate creative or expressive acts but as an expressive artefact in of itself. Magnusson (2009, 2010) argues that instruments should be idiosyncratically tied to their creators, embodying musical ideas to be explored within performance. As well as exploring a space of musical ideas, such an approach is an exploration into how action may create sound (Hamman 1999; Hahn and Bahn 2002).

We saw this perspective when we considered the composed instrument (Section 2.2.2) as well as in Chadabe's (1984) interactive composition (Section 2.2.4). However, others have criticised this approach. Jordà (2004) argues that 'too little striking music' is being created and presents a lack of original standardised instruments as a lack of 'serious evolution'. We will have the opportunity to address Jordà's concerns about DMIs in Chapter 6. At this stage, however, we intend to distinguish instruments that are created under this category as *composed instruments*, whilst those in the above categories we shall continue to refer to as *DMIs*.

When evaluated under the criteria of a conventional instruments, a composed instrument may seem somewhat lacking. However, an artist creating instruments as a part of their performative practice is unlikely to be aiming to be conventional. For example, Nicolls (2010) who mixes live electronics with an acoustic piano is more concerned with the response of an audience than of how effectively other performers might use her technology. Other examples of instruments constructed as a part of a single musical work include Hahn and Bahn's *Pikapika* (2002, see also Section 2.2.2) and Machover's technologically augmented opera *Death and the Powers* (Jessop et al. 2011).

In this light, we might see the evaluation of the composed instrument by how well it performs the role of other instruments as akin to appraising a stained glass window on the basis of how much light it allows to shine through. Consideration of utilitarian aspects independently of their artistic role misses the point of the work.

We have presented something of a spectrum of motivations behind DMI creation. Our purpose in doing so has been to observe that not all DMIs are created to serve as a channel of expression. For many, the creation of an instrument is a part of a greater composition process. This latter category we are describing as *composed instruments*, and our use of the term DMI will henceforth imply instruments that do not fall into this category.

For *composed instruments*, comparison to other instruments or a universal set of desirable features does not provide a valid—or particularly useful—evaluation. This does not mean we are unable to reliably evaluate work, establish general-purpose design principles or carry lessons forwards. However, rather than considering instruments used in a single run of performances as those that 'failed to make it', let us instead consider them as individual creative explorations into musical interaction. We evaluate them not by asking *how many people wanted to buy one?* but *how involved were the audience?*

It has been easier to consider the above distinction between DMIs and composed instruments from the position of those designing them because it is typically designers who publish papers about them both. However, it is not unreasonable to propose that audiences seeking performances involving composed instruments are engaged by the workings of the instrument as well as the music being created with it. Tanaka (2006) argued that listening to Sensorband (see Section 2.2.2) involved unravelling a puzzle of determining the difference voices and identities of the instruments. This notion is supported by a study by Gurevich and Fyans (2011) where some spectators of a new DMI were observed to be more interested in the instrument than the music being created upon it.

Having identified the composed instrument as a practice distinct from DMI design we will now consider what consequences this has on how they are designed, performed and spectated. In doing so we will seek to identify how constructing an instrument can itself be thought of as an expressive act and how we may then apply these ideas in the context of an interactive music system.

3.2 The Serendiptichord: A wearable instrument for dancers

In this section we describe the Serendiptichord (Murray-Browne et al. 2010, 2011), a wearable musical instrument for contemporary dance performance which I created in collaboration with the artist Di Mainstone as a part of this thesis. This will subsequently guide our understanding of the composed instrument in Section 3.3.

3.2.1 Collaborative origins

I was invited to collaborate with Mainstone as part of a commission she received to create an interactive sound piece to be shown at the *ACM Creativity & Cognition Conference 2009.* The project from conception to debut perfor-

Figure 3.1: Nicole Johnson of the dance company gloATL rehearsing with the Serendiptichord at the Woodruff Arts Center, Atlanta, GA.

mance took place over 12 weeks, with the first six of these conducted online through videoconferencing. The project developed further with subsequent performances described below.

Mainstone is an artist who specialises in creating wearable technology, such as *Sharewear* (Mainstone 2008), a wearable piece that was created through designing open-ended modular components and observing how others used them. Within the collaboration, Mainstone crafted the physical components, I created the software and sound design. Together we developed the concepts underlying the work and the interaction design.

Figure 3.2: Jennifer Essex wearing the Serendiptichord during the production of a short film of the work.

3.2.2 Development

The work was developed as an art-driven rather than technology-driven project: technology was chosen to realise artistic intentions rather than the work being created as a means to demonstrate technology (Candy and Edmonds 2002; Murray-Browne et al. 2013). Mainstone arrived with ideas from her previous work referencing exploratory movement and connection within public space which I related to this research as it currently stood, which referenced narrative, and its role in non-linear interactive sound works. Common ideas resonated such as narrative, exploration and space but also the instrument itself began to be characterised—unrestrainable, playful and illusive. A narrative emerged of the relationship between performer and instrument through stages of discovering the instrument, playfully exploring how it may connect to her body, becoming gradually more sinister as it begins to possess

(a)

(b)

(c)

(d)

Figure 3.3: The Serendiptichord. **(a)** With Di Mainstone. **(b)** Heidi Buehler rehearsing with Di Mainstone outside Berkeley Art Museum, CA. **(c)** Heidi Buehler performing at the *ACM Creativity & Cognition Conference 2009*. **(d)** Displayed at Kinetica Art Fair 2010, London.

and dominate her, reaching a climax whereupon she tears it off herself, and finally a return to the innocence of before as she resists its attempts to entice her once more.[1]

This narrative informed the technical constraints placed upon the instrument—such as discounting video-based sensing due to its sensitivity to different environments—how it sounded, how it looked, the anticipated means of interaction—somewhat dialogical (Figure 2.2.7) and conversational (Section 2.1.3)—and how the it would be presented to an audience. Throughout, there was a theme of serendipitous discovery, laying the foundations for an interaction design that focused on unguided exploration and learning, intuitive but unprescriptive. Thus, the Serendiptichord is not just an instrument, but also the narrative and performance that accompanies it.

3.2.3 Physical form

The instrument is made up of a headpiece module that rests on the shoulders with a 'trunk' that extends over and in front of the head, and two hand-held modules that may be attached to the headpiece or other parts of the body (see Figures 3.1, 3.2, 3.3 and 3.5). With an exterior of only wood and red leather, but a form inspired by the curvaceous nature of acoustic instruments, it is shaped to be elusive but enticing—that is, suggestive of its musical nature but uninformative about how one might go about realising this.

Three-dimensional accelerometers are embedded within each pod, behind the neck and within the trunk. Data from these is wirelessly communicated to a laptop where it is smoothed and converted into orientation data before being mapped to sound.

[1]Although to date (March 2012) all the dancers that have performed with the Serendiptichord have been female, there are no requirements on the gender of the performer. However, for readability female pronouns are used.

3.2.4 Mapping

Sound originates from a bank of *sound objects*, virtual instruments relying either on sampled or synthesised audio and a distinct rack of audio effects, each outputting audio in response to trigger events of varying velocity. Each sound object is assigned to a specific orientation of both the left pod and neck sensors, collectively referred to as the *noisemakers*. When the orientations of both noisemaker sensors reaches a certain distance from the sound object's respective assigned orientations, it receives a trigger event. This mapping is designed to make the mapping intuitive and reactive and thus easily grasped by both dancer and audience. Drawing on the embodied metaphor (Section 2.2.6) of a percussive instrument, the samples are effectively 'hit' into through rotating the noisemakers. This module of the mapping is described as the *percussive mapping model*. Further details are provided in Appendix A.

In order to allow both the dancer to introduce more structured dynamics to their performance and the composer to develop their musical ideas, the percussive mapping model is augmented with the more dialogical control of *intensity*. Sound objects are each associated a distinct continuous intensity parameter. Each object develops in a unique way when its intensity increases, as well as becoming louder and warmer.

The right pod sensor is called the *intensifier*. When it is shaken whilst the noisemaker orientations are close to that of a sound object, the object's intensity value is rapidly increased, which will then decay back to zero over time. To ensure strong perceptual feedback (see Section 2.1.1) is provided when the intensifier is shaken, the sound object is triggered each time it is 'intensified'. In addition, for a few seconds after a sound object is intensified, its audio is passed through a short delay that passes through a pitch-shifter before feeding back on itself.

In contrast to the other three sensors, output from the trunk is directly mapped to the parameters of a frequency shifting effect applied to the master channel. The trunk typically follows the motion of the dancer's neck making it an effective tool to translate expressive movement into expressive sound.

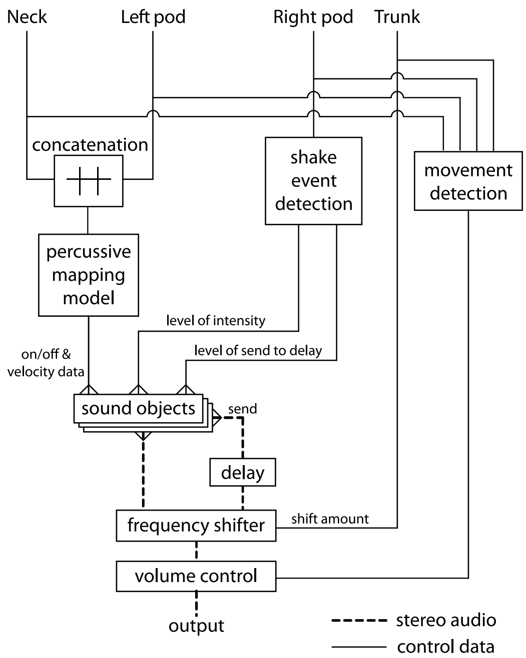


Figure 3.4: An overview of the mapping used within the Serendiptichord.

However, its physical construction causes it to gently oscillate after a sudden movement. The resulting vibrato effect provides a subtle but essential connection between the physical nature of the instrument and the audio output. As the trunk is visible to the dancer whilst the instrument is being worn, it also provides her (and the audience) with instantaneous perceptual feedback of the system's sensitivity to motion. Furthermore, through feeling the trunk follow her movement, the dancer is provided with a constant physical connection between herself and the instrument.

Both the percussive mapping model and the intensifier mapping draw on the embodied relationship (Section 2.2.6) between how much physical energy the instrument receives and how much sound it produces. However, both triggered sound objects and the intensifier delay effect may take up to a few seconds to decay to silence. In order to provide the dancer with the continuous possibility of silencing the instrument to demonstrate control, an extra 'kill switch' rapidly cuts the master volume when the instrument is still for more than a second.

An overview of the overall mapping is shown in Figure 3.4.

3.2.5 Performance

The Serendiptichord was premiered at the *ACM Creativity & Cognition Conference* in Berkeley, California on 29 October 2009 by Heidi Buehler, a contemporary dancer who had had around six hours to prepare and rehearse with the instrument. Following a warm reception, it was subsequently invited for performance at a number of other venues (see Section 1.5.2 for a full list) including Kinetica Art Fair 2010 where it was reviewed positively by New Scientist's *CultureLab* blog (Austen 2010).

Each performance is improvised over a narrative, which governs not just what kind of sounds are made but also the character of the dance, the intensity of the show and, crucially, which modules and capabilities of the instrument are revealed to the audience. Throughout its lifetime, this narrative has evolved and been refined, and the instrument adapted and improved to better support it, for example through adding the capability to load different banks of sound objects as a show progresses via backstage control. The version presented below is the current state into which it has stabilised.

Performance narrative

A typical performance will last ten minutes. For the first minute, only a dark box housing the instrument is visible. The dancer performs around the box as if magnetically drawn to it but also unaware of its contents, creating a sense of anticipation that something of significance is inside.

Once opened, just the left pod is revealed and the dancer explores how it relates to the sounds being heard. It is apparent that sounds are caused through moving this pod although the relationship between movement and sound remains unclear. The dancer emulates our limited understanding and explores this relationship. The nearly identical right pod follows and we

Figure 3.5: Nicole Johnson from the dance group gloATL with Tim Murray-Browne rehearsing with the Serendiptichord at the Woodruff Arts Center, Atlanta, GA.

might expect, as the dancer seemingly does, that it behaves in a similar fashion. But these expectations are not met: shaking it does nothing at all. Puzzled, the dancer returns to the left pod and we eventually discover that shaking it intensifies the sounds triggered by the left hand. The headpiece—perhaps the most distinctive part of the Serendiptichord—is not revealed until around a third of the way through the performance. After the headpiece has been demonstrated on its own, it is used together with the pods.

With each individual aspect of the mapping demonstrated, the limits and combined capabilities are then explored. The narrative takes a darker turn as the instrument begins to dominate and overpower the dancer. The sound and movement grows more chaotic and disturbing before we approach the climax where the dancer rips the instrument off her. Briefly, there is a recapitulation as the dancer regains herself and the sounds of the instrument returns to the earlier innocence (although audibly disfigured to some extent). But the dancer resists its allure and packs it back into its box, seals the box and leaves the stage.

The above narrative not only provides a structure for the dancer to perform over but also serves to closely control the means by which the audience's understanding of the instrument is constructed. In the following section, we will explore the consequences of this and why it is a construct particular to the composed instrument rather than the DMI.

3.3 Reflections

As a performance, the Serendiptichord fits within both the interactive dance paradigm (Section 2.2.4) and that of the composed instrument as described at the beginning of this chapter. The instrument is entwined within the described performance. This is reflected in the language of Section 3.2: we do not talk of the performance of a piece of music with the Serendiptichord but of the performance of the Serendiptichord. Attempting to evaluate its capability to fulfil a diversity of different roles as we saw in Section 3.1.1 is not helpful because that is not its purpose. But there are a number of other important consequences of this approach beyond evaluation.

3.3.1 Composing in the language of musical interaction

Most Digital Musical Instruments (DMIs) are constructed for the benefit of, and evaluated by, potential performers (e.g. Hunt et al. 2000, 2003; Poepel 2005; Arfib et al. 2003; Levitin et al. 2002). This is in line with the utilitarian motivations of Section 3.1.1 where instrument is essentially a tool to allow performers to create music. The success of the performance depends on their musical abilities as well as the extent to which the instrument served its role. As well as being the key 'stakeholder' with an interest in the quality of the instrument (O'Modhrain 2011), the performer is therefore in a privileged position to determine the impact of the instrument on the quality of the performance (Stowell 2010).

However, when the instrument is a part of the artistic contribution of a performance, the picture is different. As we saw in Section 3.1.3, the attention of the audience is on the instrument as well as the sound it produces. Therefore, the instrument designer and the performer are working together to construct a single performance of music and instrument. They are both concerned with how an audience perceives the entirety of the performance. Note that although some DMI research has considered the role of audience perception of an instrument (e.g. Schloss 2003; Fyans et al. 2010; Hsu and Sosnick 2009), it has typically done so from a utilitarian position: audience understanding needs to be established to *facilitate* the expressive performance of music (Fels et al. 2002). This is in line with a presumption among DMI creators that they belong in the Western concert tradition where audience communication has not needed much consideration (Gurevich and Fyans 2011; Small 1998).

Here, we instead have instrument and sound both as artistic contributions to a single musical performance. However, it is more than simply a double presentation. As we saw in Section 2.1.5, designing an interface is the designing of an experience. When the interface is to be used in performance, it is a vicarious experience for the audience. For example, in the Serendiptichord narrative, the performer emulated a limited understanding of the instrument, which provided the context for the audience to learn vicariously how it worked. Designing an instrument provides the framework for interaction to be investigated, explored and questioned (Hamman 1999). Therefore, performance of a composed instrument is not just the presentation of sounds, but of gestures, mappings—how those gestures relate to sounds—and interactions—which potential relationships are acted upon by the performer. In Section 2.4, we saw Small's (1998) argument that gesture is a more representational means of communication than sound. In this way, the composed instrument fulfils the role envisaged by Chadabe (2004): music as a medium of ideas as well as abstractions.

A composition is not just a collection of sounds, but a temporal arrangement of those sounds crafted to manipulate a listener's auditory perception (Section 2.4). Reasoning about composed instruments in sole terms of the musical interactions they afford (e.g. Magnusson 2010; Fiebrink et al. 2010; Schnell and Battier 2002) is akin to reasoning about a musical composition as an unordered collection of chords. Fundamental to a musical work is how these individual components are organised and the relationships between them. In this sense, musical interaction is a medium: it is not enough to simply choose interactions. You must compose them. In terms of the Serendiptichord, this means that the mapping, sound design, gestures and physical form are not enough by themselves: the narrative—both scripted and improvised—is the recipe that turns these into a musical work.

3.3.2 Perceiving a composed instrument

We saw in the previous section that the designer and performer of a composed instrument unite to present a single cohesive performance for an audience. As a result, the measure of success and the focus of development is how this audience will receive the work. Traditional NIME issues such as learnability (Section 2.2.12) and controllability (Section 2.2.11) are important, but only in pragmatic terms of how they affect the experience of the audience. For example, a number of Serendiptichord performances were with a dancer without previous experience of it and typically no more time than a day would be available to rehearse. This would be an unreasonable amount of time for even an experienced musician to learn a new instrument—but not necessarily a new piece. In comparison with NIME research, a focus on performance outcomes rather than properties of the tools used to realise it is perhaps closer to that of theatre or the performance art (e.g. Saltz 2001; Siegel 2009; Reeves et al. 2005).

In a traditional instrumental concert, there is a ‘pact’ between audience and performer establishing that what is heard should be attributed to be performer rather than machinery hidden inside the their instrument (Siegel 2009). However, in more theatrical contexts it is perfectly reasonable for technology to be scripted and controlled backstage (Saltz 2001). The Serendiptichord exploited this ambiguity by framing it as a puzzle. Whilst the dancer behaved as if she were creating the sounds, the context of the performance—her dramatic entry and unpacking of the box—implied that this control may have been an illusion. As additional aspects of the mapping are presented, further causality between gesture and sound may be perceived (see Section 2.2.10) eventually solidifying with the introduction of the trunk.

However, an audience that is unsure of what kind of performance they are spectating may be left unsure what is being exhibited for their appreciation and which parts are simply tools in service to these aspects. If left unconsidered by a designer, this can create problems—for designers of both DMIs and composed instruments. In Section 2.3.2, we saw the argument that engagement with a work relies on an audience’s ability to relate it to their understanding of the formal characteristics of a genre. A DMI perceived as a composed instrument may be seen lacking in originality or temporal development. Conversely, a composed instrument perceived as a DMI may be seen as poorly engineered technology impeding a musician’s ability to express themselves musically.

The expectations of an audience is an important aspect of musical performance that is often overlooked (Widmer et al. 2007), and we will see in subsequent chapters that this is true of interactive music too. In Chapter 6 we will consider in more detail how general expectations arising due to the performance context may affect how a work is received. However, we saw in Section 2.4.3 that evolving expectations play a fundamental role in the perception of a musical work itself. If, as argued above above, the composed instrument is most effectively thought of as a composition of musical interactions as well as sounds then the expectations of the audience needs to be considered in terms of interaction as well as sound.

For example, the second pod of the Serendiptichord is nearly identical to the first and presented immediately afterwards. As shaking the first pod resulted in sound, an expectation is established that the second will behave likewise. However, this expectation is violated when no sound is heard at all. We are left with a sense of ambiguity which, following the language of the Implication-Realization model (Section 2.4.3), we anticipate to be resolved through a subsequent realisation. In this way, we may see the perception of a composed instrument as a continually unfolding understanding of the instrument's mapping and affordances, as well as how the performer intends to make use of these. As a result, two aspects of a composed instrument performance require specific consideration: *prior knowledge* and *continuity*.

3.3.3 Prior knowledge

Slowly revealing what the Serendiptichord is forms an integral part of what makes the performance. Aspects of the instrument are revealed with musical and gestural ideas together to form the greater composition. The mapping is not explained in a pre-concert talk any more than the plot of a novel in the preface. An audience that does not know what to expect, whether they are supposed to be seeing a musician or a dancer or whether the sound is prerecorded or live is a gift rather than a hinderance. The hook of the show is to pique curiosity, raise questions and delay the answers.

In contrast to Fels et al. (2002, see also Section 2.2.10), creating an instantaneously transparent instrument is not the goal. Quite the opposite—the Serendiptichord is deliberately elusive in its design. The sensors are hidden. Whilst its shape is suggestive of musical instruments, it does not directly reference any acoustic instrument and so does not indicate visually how it might be performed nor how it might sound. Even then, it is kept sealed inside a black box until a good minute after the show has started. Everything is designed to be opaque, so that it might be communicated in a controlled way throughout the performance.

In fact, not *everything* is designed to be opaque. As we saw above in Section 3.3.2, an audience may not even consider any interactive technology is in use within a dance performance. Therefore, while it may not be clear how gesture creates sound, care is taken to communicate fairly quickly that there is indeed some meaningful relationship between gesture and sound, signalling to the audience that this is a problem for them to work out. This is similar to Edmonds et al.'s (2006) combination of initial and long-term sustainers in *Absolute_4.5* (Section 2.1.4).

The narrative of the Serendiptichord was crafted around discovery and it took advantage of an audience that knew very little about what was to come. However, this is not to say that this is a 'correct' way of creating a composed instrument. Rather, the point has been that how an instrument is perceived is dependent upon the perceiver's prior knowledge. This is a necessary consideration when composing an instrument.

Therefore, rather than being considered a static attribute of the instrument, transparency is a subjective aspect of the audience's perception that unfolds throughout the performance in a controlled manner. Likewise, communicating intention (Section 2.2.14) becomes a tool involved in this process rather than simply a means to facilitate the appreciation of skill (cf. Fyans et al. 2010). There are times when communicating a mismatch of intention and outcome is what is desired, not to allow an error to be perceived but to violate an expectation and indicate that things are more complex than they at first seemed. With the Serendiptichord we saw above an example of this with the introduction of the second pod. The dancer's response makes it clear that this is a part of the narrative rather than a fault and the whole affair is an ambiguity awaiting resolution.

It is important to note that even if nothing of a composed instrument is communicated prior to a performance, the audience will still have expectations about what is possible (given current technology) and what the norms are of musical interaction. As with traditional composition, a composed instrument is created and perceived against a set of stylistic norms.

3.3.4 Continuity

We saw in Section 2.4 that the need to balance of novel and familiar material within a musical work is analogous to the flow channel (Section 2.3.1). Continually matching challenge and skill necessitates an 'upwards path of complexity' (Blaine and Fels 2003) when learning how an interface works (Section 2.2.12) as well as when engaging with media (Section 2.3.2). It is therefore reasonable to argue that an engaging composed instrument performance is continually providing *predictive information* (Section 2.4.3) not only in terms of what we hear but in terms of how gesture creates what we hear. For this reason, presentation of a composed instrument needs to be a *continuous* process throughout the performance. It is not a precursory training session to enhance the audience's appreciation for when we move on to the 'proper' music.

In both a normal instrumental performance and that of a composed instrument, we are communicating to the audience a signal out of which increasingly complex structures emerge. Just as there is evidence of a limit to the 'cognitive bandwidth' of an instrumentalist (Levitin et al. 2002), there is a limit to how much information our audience can process at a given moment. Even if each aspect of the performance–music, gestures, revealing the mapping—made sense individually, they may not when combined simply because their overall complexity exceeds what the audience can make sense of. Many artists of composed instruments understand this intuitively. For example, Nicolls (2010) points out that a performance with complex technology should be balanced with more direct or simplified musical substance.

This notion of *complexity* is subjective to a given audience and refers to parts of the performance which are novel and to which they are paying attention. For example, the microphone introduced a number of complexities within performance (Lockheart 2003) but its role is now firmly established within the formal characteristics of many modern genres and so it does not increase the overall complexity of a performance.

3.3.5 The double bind of Digital Musical Instruments

The above observations have consequences for those designing DMIs that are not composed instruments. With an audience unfamiliar with the DMI, there is a lot for them to pick up. Many composing for DMIs recognise this and create music specifically to teach the audience about the workings of the instrument (e.g. Schedel 2002), perhaps then moving onto more complex musical ideas. Two issues arise.

Gurevich and Fyans (2011) reported spectators of a new musical instrument stating that their attention shifted during the performance from the instrument to the music. This is similar to Rokeby's (1998) argument that we stop sensing something once it has been identified. The audience jump from an completely unguided attempt to form a mental model of how the instrument works into an entirely different medium of being taken on a musical journey by the composer. As well as the potential disjunction between two different targets of attention, the audience is being asked to expend a considerable amount of 'work' in the early stages of a performance before the aesthetically rewarding material of interest may appear.

The second problem arises with repeated viewing. As mentioned above, complexity is a subjective aspect of perception. If the first part of every performance is there to communicate the instrument, then the first part is always the same. Seeing the instrument perform time and time again, possibly different performers playing pieces by different composers but each constructed to have this 'tutorial introduction' becomes equivalent to each piece opening with a tired cliché.

Therefore there arises a conflict between the utilitarian motives that see the instrument as a tool *for* expression and the artistic motives that see the instrument as an act *of* expression. The tool is created to be assimilated, forgotten—to let the music do the talking. The instrument as creative expression captures some of the limelight away from what is being heard. The tool is successful when it is ubiquitous, a formal characteristic of the genre. Until that moment it struggles because the audience is going to be figuring out how it works instead of what the music is about. The instrument as expression seeks to be idiosyncratic, original. It may fit in with some instrumental conventions to provide some conceptual landmarks from which the audience may base their understanding but ultimately it is there to grab their attention and lead them on a journey of discovering how sound and gesture relate.

In Chapter 6, we will see that this double bind has implications for the perception of skill as well as complexity.

3.4 Conclusion

Different designers of Digital Musical Instruments (DMIs) have differing motivations. Although less acknowledged, we have argued that many DMIs are created for artistic instead of utilitarian purposes. This has consequences in design and performance and we have argued that trying to do both can lead to problems. DMIs created for artistic expression we have considered as *composed instruments*.

Musical interaction is an expressive medium in and of itself. Whilst a composition may be considered as 'organised sound' (Varèse and Alcopley 1968), a composed instrument may be considered 'organised musical interaction'. Therefore, as well as choosing interactions, the order and manner in which they are presented to an audience is as much a part of the composed instrument as the mapping that defines what interactions are possible.

However, this does not mean that we are left with only aesthetic taste to guide the design and performance of composed instruments. Just as conventional music needs 'good' instruments and an understanding of how they work, those creating composed instruments can benefit from an understanding of the role perception plays in their reception. Drawing on models of musical perception, we have argued that careful consideration needs to be given what an audience expects—both in the general sense with regards to the performance context and within the performance itself. Key to this has been a consideration of how *structures*—information that allows the audience to understand and predict what will happen—emerge.

This chapter has been concerned with performance rather than interactive music. However, I have argued that an Interactive Music System (IMS) may be considered as a participatory composed instrument. By considering how musical interaction may function as an expressive medium within performance, we may now consider how the above lessons may apply within the interactive domain. In the next chapter, I introduce a more formal definition of *emerging structures* in order to model how a participant explores an IMS. Later on in Chapter 6, we will then consider in more detail issues of perceiving skill and the effects of expectations established by the performance context.

Chapter 4

Emerging Structures: A model of exploration

In this chapter we take our idea from the previous chapter of composed instruments with emerging structures of musical interaction from performance into the interactive domain. To do so, I outline a few theories from the wider field of research into how and why individuals explore and play. Combining these with the previously outlined theories of musical perception, I present a new model that more formally defines what we mean by emerging structures and how they may characterise exploration. With the model, we may then more closely examine scenarios where a participant terminates their interaction. This leads us to the proposal that captivating *interaction requires the participant to be continuously improving their understanding and potential for action—and expectant of further improvements.*

In the previous chapter, we saw that a new musical instrument that presents a novel means of creating sound may be seen as as much a part of the artistic content of a performance as the music. We are presenting not only musical ideas, but an original interpretation of how gesture may create music, an approach we described as the composed instrument. We argued that instead of having a ‘demonstration period’ to the music, the development of the mapping may be more effective as a part of the musical work itself. In this way, rather than composing the exposition, development and recapitulation of what is merely audible, we are also composing *musical interaction*. Music becomes a medium of action creating sound rather than solely a medium of sound.

In this chapter we will apply these ideas to interactive music, exploring how we might create systems where both sound and mapping are continually developing. We should be clear that our motivation is not to bring the thrill of music-making to non-musicians (cf. Blaine and Fels 2003). In line with Tanaka (2006), interactive music is created by a composer as a crafted experience of musical interaction.

What might such an experience entail? If a composed instrument embodies a diversity of musical ideas simply through its construction and mapping then we might think of interactive music as transforming this from a system explored by a performer for an audience to a system for an audience to explore by themselves. It is a composition of not only sound but of *how actions creates sound*, a relationship to be experienced by the audience actively.

However, whilst performing with the composed instrument allowed us to carefully control the order and manner in which this exploration took place, letting the audience take control presents difficulties. Benford and Giannachi (2008) describes this issue in terms of a participant’s *trajectory*, the ‘journey’ taken as they explore a system. The imagined *canonical trajectory* of the designer is often rather different from the actual trajectory of the user (Benford et al. 2009). In particular, Rokeby (2011) observed that people at interactive installations are often focused on ‘working it out’ and not really experiencing it as the artist had in mind.

So what goes wrong when things become interactive? Why do we end up with interactions of 5–10 seconds when the audience is left to explore on their own devices (Section 1.1) but seem to be able to engage audiences for significantly longer when they are watching somebody else explore the system? Why does the audience apparently choose to 'figure it out' and get quickly bored once they have done so, but gain vicarious pleasure from watching another person drag out the exploration?

To answer these questions, we develop in this chapter a model of *emerging structures* (ES) that will allow us to reason about how individuals go about exploring an interactive system. From here we may consider how to make interactive music *more like music* and less like a puzzle to be 'worked out'. We propose that through doing so the experience will be made more captivating.

The model presented in this chapter is fairly abstract in nature. However, the understanding it provides us will allow us to form a more practical design principle in the next chapter.

4.1 Theories of exploration, play and curiosity

Before we can develop the model, we need to review some existing research into how we explore and what motivates us to do so. In particular, we would like to clarify what exactly the distinction is between 'figuring something out' and 'exploring something'. This will allow us to untangle why the former may lead to a less than captivating experience.

4.1.1 Investigative and diversive exploration

Costello (2009) considers this distinction in terms of *play*, for which she adopts Zimmerman's (2004) definition of *free movement within a more rigid structure*. Drawing together a range of theories, she concludes that play involves oscillating between two types of behaviour, *investigative* exploration ('what does this object do?') and *diversive* exploration ('what can I do with this object?').[1] We will be relying on this distinction a number of times in

[1] Costello's terms are adopted from Hutt (1985), which are based on Berlyne's (1966) distinction between *specific* and *diversive* exploration.

this chapter. For Costello, investigative exploration is the establishment of Zimmerman's *rigid structure* in which diversive exploration (*free movement*) may take place. The *pleasure* of play is to be had in diversive exploration. Boredom then triggers a return to investigative exploration (Costello 2007). We may also hypothesise that a serendipitous discovery could cause this transition. However, far from being a state of low arousal, boredom arises because a learning process has reduced the perceived complexity of the present activity (Rauterberg 1995). Through this desire for greater complexity to relieve understimulation, it plays a fundamental role in establishing curiosity (van Aart et al. 2010).

In this way, exploration is seen as an oscillating process between investigative and diversive exploration, with the desire to apply the fruits of our labour motivating a transition from the former to the latter, and boredom with what we can do motivating a transition back to investigative exploration. However, boredom (along with confusion) also seems to be a common reason given by those who fail to engage with systems (e.g. Costello 2009, p. 140). So what goes wrong?

4.1.2 Exhausting the (perceived) possibilities

Insight into why boredom might trigger differing responses arises in Gurevich et al.'s (2010) study investigating highly constrained musical interfaces (see Section 2.2.2).

Gurevich et al. identified a spectrum of exploratory approach to describe their participants' behaviour. They describe at one extreme of this spectrum participants following a *vertical* approach as identifying a single playing technique and attempting to exhaust its entire potential until something new emerged, whilst at the other end those following a *horizontal* approach are described as attempting to find as many playing techniques and musical possibilities as possible (Gurevich et al. 2010).

We might respectively see the horizontal and vertical approaches as being predominantly investigative and diversive. Somewhat paradoxically the participants most likely to report that they had 'exhausted the possibilities' of the instrument were those in the middle of the spectrum, with those on either end believing that there was more to discover. However, what did correlate with this notion was the participant's perception of their own skill with the instrument. Those who saw more possibilities perceived themselves as less skilful. It seems that it is this *believed potential for further mastery* that inspired some participants to continue—as before, either through uncovering new inputs or through developing a deeper understanding of those that have been explored. This is similar to the theory of flow (Section 2.3.1), except rather than simply balancing challenge and skill, here we see that it is the ability to *identify* challenges and the *possibility* of skill development that maintains interest.

A participant disengaging due to a sense of having exhausted the possibilities of what they could do was also noted by Costello et al. (2005), who observed that right before this moment, participants repeated actions that they had previously performed at their most intense period of engagement (see Section 2.1.6).

The problem appears not to be that some are simply more curious than others but that some are able to see more possibilities to investigate. In order to understand why these individual differences arise—and therefore how we may compensate for them—we turn finally to some research in cognitive psychology that investigates what *motivates* us to explore.

4.1.3 Curiosity

Loewenstein (1994) describes curiosity as the desire to close a perceived knowledge gap under his Feeling-Of-Knowing hypothesis. The 'magnitude' of this gap is determined by how close an individual feels they are to the knowledge. In a series of studies, Loewenstein found that as the feeling-of-knowledge gets closer (towards, e.g., the tip-of-the-tongue phenomenon), curiosity increases. Crucial is the sense that there is something *new* to be found out. A sense of *certainty* reduces curiosity (Loewenstein 1994).

Litman and Jimerson (2004) expand on this with their Interest/Deprivation categorisation, which identifies two types of curiosity. Curiosity as a feeling of deprivation (CFD) is caused by a perceived knowledge gap. Curiosity as a feeling of interest (CFI), on the other hand, is associated with a more hedonic, pleasure-seeking and aesthetic interest. Because CFD reflects an unsatisfied need, it is argued to be a more intense desire than CFI (Litman 2005). However, further studies suggest that CFI stimulates more positive affect, diversive exploration and mastery-oriented learning. CFD is focused on reducing uncertainty and is more closely associated with performance-oriented goals, i.e. goals in which an outcome may be deemed a success of failure. In particular, satisfying CFD leads to a sense of *closure*, i.e. a termination of the curiosity (Litman 2008). As the feeling-of-knowledge gets closer and the anticipation of an answer draws near, tension increases. If the answer is not then forthcoming, frustration results.

Litman (2008, 2010) has highlighted the role that individual differences may play in determining which type of curiosity predominates; in particular, an individual's tolerance for ambiguity (which reduces the need for closure). However, the interest/deprivation distinction demonstrates how participants who are primarily motivating in 'solving' an installation might have something of a shorter and less hedonic experience than those with less specific goals.

4.1.4 Artificial curiosity

Curiosity has been modelled as a means to guide the autonomous exploratory learning of robots, a field described as 'developmental robotics' due to its inspiration from the developmental processes of early childhood.

Oudeyer et al. (2007) model curiosity as a drive that allows learning to be maximised. Their model, Intelligent Adaptive Curiosity, is defined in terms of the sensorimotor context—a robot's sensory input of its environment and estimation of the state of its actuators (Vetter and Wolpert 2000)—and actions that the robot may perform across three layers. The first layer is a classic machine learning system that predicts a change in context based on an action-context pair, the second a *meta-machine learner* that predicts the error of the first layer over action-context pairs. In the third layer, actions are chosen by a *knowledge-gain assessor* that draws upon the second layer to maximize the decrease in the mean error of the first layer over the entire history of sensorimotor contexts.

Schmidhuber (2009, 2010) also considers a computational model of curiosity. Rather than minimising predicted error over situated action-context pairs, his model seeks to maximise the compressibility of the history of observations. Compressibility is determined as a subjective measure determined by both the abilities of the agent and their prior observation history. Under this model, an agent in the absence of external rewards will focus attention on areas that are expected to potentially yield further improvements in compression over *prior* observations. Such improvements will be found through partially novel observations that highlight structural similarities across these observations, which Schmidhuber (2009) argues lead to the Wundt curve peak (see Section 2.4.3).

The two models are similar as compressibility is implied by predictability. However, with a single parameter being maximised, neither seem to satisfactorily encapsulate what causes the individual differences identified in the previous section. We may note that Litman's (2010) evidence for individual differences within the Interest/Deprivation categorisation was collected from adults whereas the models here were inspired by the developmental processes within early childhood.

Whilst the above approaches place a value on forming an understanding of the effect of one's actions, they do not place a value on the efficacy of these actions in allowing the agent to affect and manipulate their environment. In the language of Section 4.1.1, we might describe their focus as being solely investigative rather than diversive. However, in recent work by Baranes and Oudeyer (2012) exploration is guided in terms of a task space rather than the sensorimotor space. Their Self-Adaptive Goal Generation architecture maximises an agent's *competency* rather than its knowledge by forming goals within a task space. In this manner, their curiosity drive focuses the agent more towards discovering how to perform multiple tasks rather than multiple ways to perform the same task.

The model we present in this chapter will consider both knowledge and competency. Our work is motivated more by the desire to describe the exploration of interactive music by an adult and its aesthetic and motivational consequences, rather than the development of basic sensorimotor abilities by a robot or child. For this reason, the Emerging Structures model presented later in this chapter draws on language more oriented around the musical-perceptual models described in Section 2.4.3 rather than that from the field of artificial curiosity.

We have seen a range of theories in this section approaching the question of how and why we explore, each addressing the topic from a different angle. Along the way we have defined *investigative* and *diversive* exploration (Section 4.1.1), *horizontal* and *vertical* exploration (Section 4.1.2) and curiosity as a feeling of *deprivation* (CFD) or *interest* (CFI) (Section 4.1.3). We might argue that those 'figuring out' the system are performing investigative exploration rather than diversive, motivated by a perceived knowledge gap rather than an interest in what they might do with it. Once the system has been worked out, it offers little to discover and thus the possibilities are exhausted.

However, although Gurevich et al. (2010) demonstrated that constraints may spark a diversity of approaches, we have not established the role that the design of the system plays in encouraging specific types of exploration. In particular, how might a system lead a participant into the perception that there is no more to discover?

Whilst the above theories all seem to provide a piece of the puzzle, it can be difficult to see how they relate with each other. For example, we saw that investigative and diversive exploration were perhaps related to horizontal and vertical. But without a common means of expressing these observations it is difficult to see exactly how, and what consequences we might draw from this. In the next section, we present the Emerging Structures (ES) model of exploration. It is a descriptive model that will allow us to unite the above theories and address notions of perceived possibilities with more rigour. With this model, we will then be able to derive a number of predictions in Sections 4.4 and 4.6, some of which we will carry further within this thesis.

4.2 Emerging Structures: A model of exploration

In Section 2.4.3, we saw theories of musical perception in which a listener seeks to construct a predictive model by which to perceive structural consistencies in what is heard. The composer then crafts the expectation, anticipation, affirmations and surprises into their music in a manner that both allows and manipulates a continuous expansion of this model. We then saw in Chapter 3 how with the composed instrument this model might be applied to interactions with sound as well as to sound itself. In this section, we develop a new model of how we explore Interactive Music Systems (IMSs) based around the notion of *emerging structures* (ES). Our approach extends the expectation-based model of perception into the field of interactive music. As we shall see, making things interactive presents a unique challenge.

The choice of the term *structure* within the model draws upon Minsky's (1981) analogy of conceptually constructing a representation of a physical scene from what we see (see Section 2.4.3). We will be applying this perspective to a participant piecing together the mapping of an IMS based on its behaviour. As with Minsky's analogy, we are developing an understanding of something that (in its simplest form) is static through a dynamic process. In Minsky's case this was a scene through seeing; in ours it is a mapping through interacting.

4.2.1 Mental models of mappings

We saw in Figure 2.1.1 that users of a system establish a mental model of how it works. A key component of this is the *perceived mapping*, the user's expectations about what outputs may result from a given input. As argued in Section 1.3, as we are focusing on the *experience* of interacting with an IMS then it is these perceived attributes of a system that are of importance rather than what is 'actually' going on. Throughout Chapter 2, we saw that although interactive systems were often discussed objectively in terms of their behaviour, these were often implanted with subjective notions of how we perceive the system. For example, Cornock and Edmonds's (1973) described a category of systems that included an invisible agent modifying the original specifications of a work (Section 2.1.2). We make sense of a system by constructing a perceived mapping of what we can do with a system. When an earlier perceived mapping changes, it makes sense to think of a module within the system modifying its behaviour. However, on purely objective terms, every response of the system has been in accordance with the mapping of input to output programmed within it.

In the emerging structures (ES) model, we will consider mappings as *perceived* entities: expectations the system has established with the participant about how their input will affect its output. The perceived mapping is the participant's working estimate of the system mapping. Therefore, the only limitation of the complexity of a mental model lies with the cognitive abilities of the subject. Their model of the mapping may be dynamic and non-linear. It will be informed by the system's prior behaviour as well as any other sources of information the participant may have, for example experience with similar systems in the past, instructions for use provided by the artist or having witnessed another participant use the system.

Note that a mental model may be dynamic in the sense that it is changing over the course of an interaction as well as being a model of a dynamic mapping. For clarity we will describe

- the *mental model* as the participant's cognitive representation of the

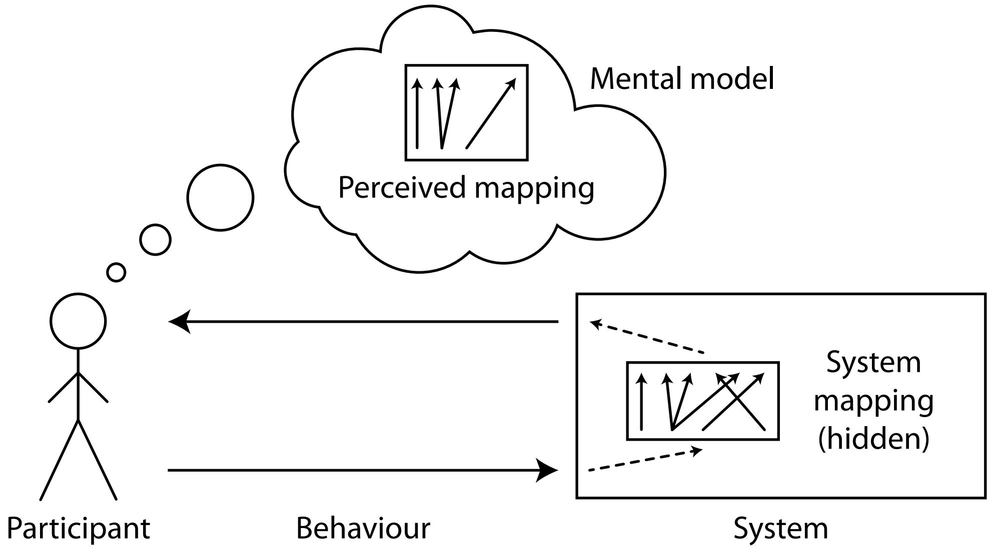


Figure 4.1: The participant's mental model estimates a perceived mapping. This is determined only by the behaviour of the system, which is itself determined by the system mapping.

system,

- the *behaviour* of the system as the input and output of the system that is perceived by the participant,
- the *perceived mapping* as the expected behaviour of the system as represented by the participant's mental model, and
- the *system mapping* as the programmed rules of what output results from an input and so determining the behaviour of the system.

For example, a participant may have an unchanging understanding of a system with a dynamically evolving system mapping. In this situation, the mental model would be static whilst the perceived mapping would be dynamic. These four definitions are illustrated in Figure 4.1.

Fundamental to what follows is the distinction between perceived and system mappings. From the participant's perspective, the system mapping remains invisible. Even if the participant is the programmer of the system mapping, what they are left with is their understanding of its behaviour—the perceived mapping.

Our focus will be on how the behaviour of the system constructs a perceived mapping during their interaction. As acknowledged above, a participant may also learn from other sources. However, we will consider this as *prior knowledge* that was formed before the participant interacts with the system. We will consider only interactions where the participant has no prior knowledge of this particular system, although further extensions to the model may include these aspects.

Note that we are always considering a mapping to be a *perceptual* quality of the system that may be constructed by its behaviour. The system mapping is only relevant insofar as it affects the behaviour of the system.

An input may be any gesture of the participant that is sensed by the system. As our focus is on perceived mappings rather than system mappings, we will not need to be more specific than this. Different participants using different systems will have different notions of specificity as to what an input is (e.g. moving quickly, pushing a button, standing in a particular location). However we consider the time a gesture is performed as encapsulated by the notion of an input. A button pressed at a specific moment may lead to expectations about what will happen when that button is pressed at a later moment but that later press would be considered a different input.

4.2.2 Consistency

As a participant interacts with a system, we will describe the behaviour of the system as being *perceptually consistent* at a given moment and input if its output is as predicted by the mental model the user has established up until this point. The participant is an individual from a set of potential participants with different ways of thinking. A behaviour may appear consistent to one participant but not to another. It will at times be useful to think of an 'ideal' participant with a maximal ability to construct an accurate mental model based on the behaviour of the system. Behaviour that would be considered consistent to this ideal participant we will refer to as *potentially consistent*.

Note that a mapping may incorporate non-determinism and still be perceived as consistent provided the behaviour of the system is within the range of outputs the user's mental model predicts (what O'Modhrain (2011) describes as 'reliably unpredictable'). Consistency is therefore a distinct concept from predictability or controllability.

4.2.3 Structures: The units of a mental model

An advantage of distinguishing between the system mapping and the participant's mental model is that we may now examine the nature of how the perceived mapping is arrived at. One of the key conclusions from the previous chapter was that controlling what aspects of the system are revealed when is instrumental in how engaging a system is.

To do so, we define the *certainty* of the perceived mapping as its perceived ability to predict the future behaviour of the system. We have said that the complexity of a perceived mapping is limited only by the cognitive abilities of the participant. As such we shall not attempt to define how such a mapping is represented by a mental model. However, it is fair to assume that consistent behaviour has the potential to improve the participant's perceived ability to predict future behaviour. It will therefore be helpful to reason in terms of 'pieces of knowledge' that when added to the participant's mental model increase the certainty of the perceived mapping. We will refer to such pieces as *structures* and the event of their addition to the mental model as their *emergence*. We may think of the emergence of a structure as the recognition of a pattern that may be extrapolated to predict the outcome of future inputs. In information theoretic terms, the emergence of a structure is equivalent to the observation of a redundancy within a signal.

As before, we may imagine some participants recognising patterns that others do not. Therefore, we may think of a *potential structure* as a structure that would emerge for the 'ideal participant'. For a given moment, we may then define a *maximal potential mental model* as the set of all potential structures.

If the behaviour of the system is inconsistent with the predictions of the perceived mapping then certainty may decrease and we may also imagine a structure being disregarded as erroneous. However, if the participant believes that the system mapping is consistent then this may be received as a challenge to find a *greater structure*—a structure that reduces the overall uncertainty in spite of the earlier loss.

The above describes the basic frame of the ES model: the participant constructs a mental model that defines a perceived mapping allowing them to predict the behaviour of the system. As the behaviour of the system is observed, structures are added and removed to the model with the respective effect of increasing or decreasing the uncertainty in their model. We may therefore form a basic notion of exploring a system as a process of providing inputs that produce behaviour that aids this process.

4.2.4 Size, uncertainty and capacity

However, in Section 4.1.2 we saw two distinct ways in which a participant may improve their ability to use a system: horizontal and vertical inputs which we may think of respectively as finding new inputs or learning more about the found inputs. From this we may see that of the inputs that may potentially affect the behaviour of the system, the participant may only be aware of a subset, which we refer to as the *exposed* inputs. Note that an input may be exposed without being performed. Exposing a gesture as an input simply means forming a belief that it will influence the system's output.

When a previously unexposed input is exposed, the perceived mapping will need to be amended to accommodate it. Therefore, as well as thinking of the uncertainty of the perceived mapping, we may think of its *size*. Together, the size and certainty of the perceived mapping determine the control the participant has over the output, which we will describe as the perceived mapping's *capacity for action* (or *capacity* for short).

In this way, we may think of horizontal exploration as seeking to increase the size of the perceived mapping and vertical exploration as seeking to increase its certainty. Both of these, however, increase the capacity for action. But we described these two approaches as being the extremes of a spectrum rather than distinct modes of operation. In order to identify this continuum, we need to be able to quantify structures in terms of the extent to which they decrease uncertainty and the range of inputs over which they do so. Therefore we define a structure as being more *global* if it reduces the total uncertainty over a larger set of inputs and as being more *deep* if it reduces more uncertainty over an equivalently sized set of inputs. Correspondingly, we may talk of structures as being *local* or *shallow*. Structures that are deeper or more global both lead to a greater increase in the capacity. We may therefore combine these measures and consider the *strength* of a structure as the extent to which it increases the capacity.

4.2.5 The perceived potential for further exploration

At this stage, we may argue that exploration is driven by the desire to increase the participant's capacity for action. However, in the box instrument study (Section 4.1.2), the indication from those who felt they had nothing more to gain from the instrument was not that this desire had stopped but that they had exhausted the possibilities to do so. This suggests that what turns this desire into motivation is a sense of potential—the expectation that potential structures or unexposed inputs are available.

The former of these two is analogous to the *expected predictive information rate* that we saw in Section 2.4.3). However, rather than being a single measure, such an expectation is formed over a subset of the exposed inputs. We shall describe it as an *implication*. The event in which expected predictive information arrives we will describe as a *realisation*, the case where this expectation is not met as a *denial*.

The other way in which the participant may expect their capacity to be increased is by discovering unexposed inputs. We describe this as their *encounter entropy* (a term borrowed from Arfib et al. (2003)).

Combining the encounter entropy with the implications that are open over the exposed input space defines the total expected increase in capacity that may be obtained, which we will describe as the *anticipated capacity*.

This completes the definition of the Emerging Structures model of exploration. At present, it simply describes an interaction rather than offering directly testable predictions. However, it provides a unifying language in which to apply some of the aforementioned research of Section 4.1 and Chapter 2, which we will do so in Section 4.4.

4.2.6 Summary

The Emerging Structures (ES) model

A **structure** *is defined as a unit of knowledge that facilitates the prediction of the output that arises over a set of inputs. The size of this set of inputs determines how* **global** *the structure is; the strength of the prediction determines how* **deep** *it is. Through perceiving the behaviour of the system, structures emerge to form the participant's mental model.*

The mental model defines the following.

The exposed input space: *a set of inputs which the participant believes may result in output*

A perceived mapping: *a probability distribution of outputs given an exposed input. From the perceived mapping we derive*

- **Uncertainty:** *the uncertainty of the outputs averaged over the exposed input space.*

Encounter entropy: *the participant's estimation of the extent to which unexposed inputs remain.*

A set of implications: *expectations of a reduction in uncertainty that will arise if a specific input is performed. Implications remain* **open**

whilst this expectation exists. Once closed they have been **realised** *if the expected reduction happened or* **denied** *if it is not.*

From this we may derive the following.

The capacity for action: *the extent to which the participant may reliably produce outputs. This is a combination of the size and uncertainty of the exposed inputs.*

The expected capacity gain: *the participant's estimate of how much further their capacity may increase. This is derived by combining encounter entropy and the total expected reductions in uncertainty defined by the implication set.*

4.3 Example: An interactive sound installation

Before we proceed to apply the above model to understand the different experiences that may arise through interacting with an Interactive Music System (IMS), we will first provide an imagined example to illustrate how the different aspects of the model work together. Our imaginary participant will be called Alice.

This example is simply to aid an understanding of how the above definitions relate to practical aspects of an interaction with an IMS. After we have derived some predictions in Section 4.4, we will provide a more concrete example in Section 4.7 when we apply the model to understand the audience experience of the Serendiptichord.

Alice enters a room within a gallery unaware that it contains an IMS. At this stage the perceived mapping is empty and encounter entropy is negligible.

She is made aware by reading a description on the wall that an IMS is installed within the room but there are no details of how it works. At this stage, as she has exposed no inputs the perceived mapping is empty but as she expects to do so encounter entropy is now high. With no instructions, Alice's expectation of what gestures might be an input are only guided by prior knowledge of similar systems and what she can see in the room. There are no buttons but there are cameras installed in the corners.

Alice tries random gestures: walking through the room, stamping her foot. Eventually she hears something after waving her arm above her head. The description that an IMS was installed led Alice to believe that there would be at least one input that would create an output. However, having found one she is now less sure if there may be further unexposed inputs. Therefore encounter entropy is now slightly lower. Alice suspects that repeating this input will cause the same output and that similar (but as yet unperformed) hand waves might cause similar but distinct sounds. However, having only provided this input once, Alice is now in a position to find out more about it through experimentation. This is an *implication*—the expectation of predictive information associated with a set of inputs (in this case, hand waves).

Alice tries similar hand waves through which she discovers that the height of her hand determines the pitch of the sound and the speed its the amplitude. This discovery is described in the model as the *emergence of a structure*. The addition of this structure to her mental model increases the certainty of Alice's perceived mapping. As a result, we would describe the implication as having been *realised*.

By chance, she claps their hands together whilst doing so and hears a bell ringing. This implies that there may well be further inputs to the system increasing encounter entropy. Exposing this input means that her perceived mapping has grown in size. The particular input is likely to be associated with moderate certainty within her mapping as she has only provided it once. As a result the overall certainty of her mapping may have decreased but as the previous structures remain in her mental model, her *capacity for action* has only increased. Once again, exposing this input has created the implication that further structures may emerge through exploring similar inputs.

Alice repeats this gesture expecting to realise this implication as she did before. However she instead hears something completely unexpected—a clap of thunder. She was expecting to have increased her certainty over this area of the input space but instead it has decreased. She suspects that the output may be determined by some attribute of the clap (i.e. clapping is actually a set of inputs rather than a single input). Although this was a surprise, the *implication* remains open as there is still the perception that a potential structure exists. She tries a few more times and hears each time a different sound. At this stage, although there remains uncertainty over the exact output associated with the input, she does not believe that it will become any clearer. The implication has thus been *denied*. As such there is no longer the belief of further structures and she therefore accepts that this is the limit of her control over this aspect of the system.

In fact, the choice of sound is not random at all, but determined by the velocity of the hands when they clap passed through a complex system mapping to select a sample. However, Alice never knows this so it is *effectively* random from her perspective. This is an example of a potential structure that is rarely established—perhaps only by participants with prior knowledge of it (e.g. the designer).

Alice returns to using their arms and wanders around the room swinging them about, losing herself as she finds more subtleties (i.e. deeper structures) to this particular input: it turns out that suddenly stopping the arms whilst swinging causes an interesting timbral effect. However, despite providing a few more random gestures along the way—swinging her arms by her hips, lifting a leg up—no further unexposed inputs become apparent and so encounter entropy drops.

She is now starting to create a piece of music with her arms and goes to add the random noise associated with a clap, only to find that now nothing happens when they clap. This is perceived as inconsistent: although she had little certainty over what would be heard when clapping, it was at least established that she would hear *something*. She tries a few more times and in different ways but still with no luck, although the arm swinging still works as before.

Alice is now slightly confused. Her capacity for action has dropped by this loss in the little certainty she had regarding the clap. She wanders around the room attempting to recreate the clapping sound and eventually does so—it turns out that it only created a sound because she was stood in a specific location within the room. At this stage the clapping inconsistency has been resolved by a greater structure: the recognition that the location in the room distinguishes between otherwise identical inputs. Although Alice's capacity for action is back to where it was previously, this recognition brings with it an increase in encounter entropy. All of the random but unsuccessful attempts at exposing inputs may have simply been unsuccessful due to Alice's location in the room.

Alice again tries different random gestures in different parts of the room but does not find anything new. Encounter entropy drops again.

She returns to exploring the clap and arm swinging movements, trying to create an original sound with them. Eventually though, she finds that she is not finding anything new. With low encounter entropy and a levelling out of uncertainty over the exposed inputs, there is no longer any perceived potential to increase her capacity for action. She decides that she has had all of the fun that she shall with this particular installation and leaves.

4.4 What makes for captivating interaction?

In this section, we will consider how the ES model may assist us in understanding what makes for a captivating interaction with an IMS by drawing together the above theories of playful exploration (Section 4.1) with the models of musical perception described in Section 2.4.3.

4.4.1 Anticipation

We use the term *anticipation* to describe the expectation of a desired event. Within the ES model, we only use the term to describe events regarding the state of the mental model, such as the emergence of a structure or the exposure of an input.

The choice of the terms implication, realisation and closure in the model are intended to draw a parallel with the Implication-Realization (I-R) model (Section 2.4.3). As in the I-R model, implications are expectations regarding the outcome of a future event that have been established by prior events. Implications create anticipation of their closure which causes tension while they remain open. With the ES model it is the participant who determines when this event might happen. Therefore, we may propose

Prediction 1 *Delaying a participant from performing the input necessary to close an implication creates anticipation of (implying desire for) its closure.*

4.4.2 Sustaining interaction

Before considering what it means for an interaction to be captivating, we may ask the more basic question of what determines whether they continue to interact at all. In other words, what causes an interaction to be *sustained*?

The Feeling-of-Knowledge hypothesis (Section 4.1.3) argued that the closer the perceived 'knowledge gap', which in the ES model corresponds to the confidence of the expected decrease in uncertainty, the greater the motivation to acquire it. Therefore, we may propose

Prediction 2 *Participants will be motivated to provide inputs that close the strongest implications (i.e. inputs with the highest expected reduction in uncertainty).*

Investigative and diversive exploration (Section 4.1.1) were described as respectively identifying what an object does and what it allows the participant to do. In the ES model, investigative exploration relates to the seeking of unexposed inputs; diversive to the seeking of structures over the exposed input space. We saw in Section 4.1.1 the proposal that exploration involves oscillating between these two states with boredom triggering a switch from diversive to investigative. However, we also saw that boredom with diversive exploration may cause the participant to disengage instead (recall from Section 2.1.6 that *disengagement* describes the period in which the participant terminates their interaction). Combining this with Prediction 2 we may propose

Prediction 3 *There is a minimum strength necessary of an implication for the participant to be motivated to close it.*

We may argue, in line with the observations of Section 4.1.2, that the determiner of which of these two possibilities arises is their perception of whether they are likely to be successful in finding unexposed inputs—i.e. encounter entropy. This gives us

Prediction 4 *If the participant is not motivated to close any implications then they will either seek to discover unexposed inputs (investigative exploration) if encounter entropy is sufficiently high, or terminate their interaction (disengagement) if it is not.*

Combining Predictions 3 and 4, we may reason that the common determiner of whether a participant continues is the expected capacity gain which combines both encounter entropy and the overall strength of the implication set. Therefore, we may assume that this capacity gain is anticipated by the participant,

Prediction 5 *Participants are motivated to continue using a system by anticipation of capacity gain.*

leading to

Prediction 6 *Systems whose behaviour creates anticipation of capacity gain in the participant will tend to sustain interaction more than those that do not.*

As well as being realised, implications can be closed through denial if the anticipated reduction in uncertainty does not arrive but expectation of it doing so subsides. Therefore, boredom may arise both through a sense that all potential structures have emerged or that no potential structures will emerge. This latter case we may relate to Costello's (2009) observations of participants who found themselves unable to understand a work and disengaged frustrated and confused. From this we may argue that the denial of an implication causes frustration in the participant. Furthermore, an interaction in which implications tend to be denied rather than realised may lead the participant to infer that future implications will also be denied. This reduces the expectation of reductions in uncertainty and consequently any anticipated gains in capacity. From this we propose

Prediction 7 *As the total strength of the implications that have been denied increases in proportion to that of those that have been realised, a participant will become more likely to disengage due to frustrated confusion.*

Prediction 8 *Systems whose behaviour creates implications but tend to deny rather than realise them will tend to trigger disengagement due to frustrated confusion.*

In other words, a participant who believed they will be able to understand a system but subsequently finds no structures emerge during the interaction will give up frustrated and confused.

These two types of disengagement, due to a lack of open implications or due to an overabundance of denials are both situations where the system's behaviour has failed to create anticipation of capacity gain. They are analogous to the two conditions minimising the average predictive information rate (Section 2.4.3)—completely predictable and completely random—as well as the cases described by the two sides of the flow channel (Section 2.3.1)—too easy and too difficult.

4.4.3 Engaging interaction

As suggested by the CFD/CFI distinction (Section 4.1.3), a system that sustains interaction by provoking curiosity may not necessarily enjoyable or engaging. Costello (2009) argue (as does Hutt (1985)) that investigative exploration is the 'work' that allows the pleasure of diversive exploration to happen. In this way, we may see investigative exploration as the 'accumulation' of implications in order to experience the pleasure of realising them in diversive exploration. From this we may argue

Prediction 9 *Participants tend to be more engaged when structures are emerging.*

from which we may propose

Prediction 10 *An interaction in which structures are continually emerging will be more engaging than one in which they are not.*

These predictions are similar to that of the information dynamics model of musical perception (Section 2.4.3): we are engaged by information that aids our ability to predict the future. However, we are defining a model of *action* as well as perception. Recall Costello's description of the motivation behind diversive exploration: *what can I do with this object?* Likewise, in the theory of flow (Section 2.3.1), it was the steady improvement of *ability* that provided enjoyment. For this reason, we argue that Prediction 9 is a special case of a more general preference for an increase in the capacity for action:

Prediction 11 *Participants tend to be more engaged when their capacity for action is increasing.*

From this we propose

Prediction 12 *An interaction in which the capacity for action is continually increasing will be more engaging than one in which it is not.*

4.4.4 Captivating interaction

Predictions 6 and 12 provide two ways in which the process of constructing a mental model may affect the experience of a participant: *actual capacity gain* determining how engaging a system is, and *anticipated capacity gain* determining its ability to command the participant's attention and provoke continued use. Combining these two measures we define a *captivating* experience as being both engaging and demanding of continued attention. From this we derive

Prediction 13 *A system whose behaviour causes continual gains of capacity and anticipation of further gains will be more captivating than one that does not.*

This is the key prediction of the ES model. In general terms, interacting with a system will be captivating if the participant is continually improving in their ability to do things—either through finding new inputs or through improving their command over those they have already found—and anticipating further improvements. In these terms it is similar to the theory of flow (Section 2.3.1) although more specific in how an improvement in ability arrives and with a greater focus on *anticipated* improvements as well as actual ones.

Whilst this prediction is difficult to test directly as capacity and anticipation are defined subjectively, in Section 4.6 we will consider how differences in systems and participants affect these factors, which will allow us to derive a more readily testable design principle in the next chapter. First, however, we will consider a concrete example illustrating these predictions.

4.5 Example: The keyboard synthesiser

Whilst the example of Section 4.6 demonstrated how the ES model may be applied to an interactive sound installation, it may also be applied to other types of IMS such as a musical instrument. Here, we will provide a simple example of a keyboard synthesiser. In this example, our participant will be called Bob and we will assume he is using for the first time a simple keyboard with a number of unmarked buttons that select different output voices.

Bob has experience of the piano but none of a keyboard synthesiser. Therefore he begins the interaction with the prior expectation that keys are mapped to pitch in the familiar manner and that pressing harder leads to louder notes. As he begins playing, these expectations are met. He then observes some unmarked buttons on the keyboard. These are *exposed inputs*; Bob expects them to influence the keyboard's output but is uncertain how they will do so. The creates an *implication* over these inputs: the expectation that *uncertainty* will decrease if a specific input is performed.

The ES model predicts that whilst this implication remains open it creates anticipation of an increase in Bob's *capacity for action*, which makes it more likely that the interaction will be *sustained*. In other words, having made this observation Bob is more likely to continue using the keyboard than had he not.

Now, suppose Bob presses a button but there is no perceivable consequence to the output of the keyboard. The implication is *realised* because uncertainty over inputs involving the button has decreased. However, Bob's perceived capacity for action remains unaltered. We may also propose that Bob's estimation of the likelihood of the other buttons also having no perceivable consequence increases. Whilst this forms a further implication (Bob can easily verify if this hypothesis is correct), the *expected capacity gain* arising from the buttons decreases as they no longer offer the potential to increase the range of outputs he may produce. Therefore, the ES model predicts that whilst Bob will be engaged whilst exploring these buttons (and indeed likely driven to try them all to make certain that they are of no consequence), they will not increase the probability that his interaction is sustained.

Consider on the other hand a scenario where the button changes the output sound of the keyboard from a piano to a church organ, an instrument whose functioning is also familiar to Bob. As he subsequently plays the keyboard the uncertainty raised when the button was exposed quickly reduces. With this new sound, how the exposed inputs may now be combined to produce new outputs creates implications driving Bob to continue exploring. Here, capacity increases not through the exposure of inputs but through the emergence of structures, and the expectation that this will continue to happen. The ES model predicts that as a result of these two conditions, Bob will be more likely both to be engaged and to sustain the interaction, respectively—the combination of which we defined as *captivation*.

At this stage, Bob has formed the expectation that each button on the keyboard will lead to it functioning like a keyboard but sounding like a different familiar instrument. This creates an implication: Bob expects that by pressing the untried button followed by a single key on the keyboard, uncertainty over a large range of inputs—how the remainder of the keys function under that input—will be reduced. However, on pressing a second button instead of hearing a different instrumental timbre Bob instead find that the pitch associated with each key now follows a 19 note equally tempered scale rather than the familiar 12. This is an example of an implication being *denied*: the expectation of a decrease in uncertainty has been lost without the decrease being realised. The model predicts that for Bob to now be engaged by this unfamiliar scale, structures need to emerge through his ability to predict the output of inputs and so exert reliable control over the output. If structures are not emerging, the ES model predicts that Bob is unlikely to be engaged although his interaction is more likely to be sustained if he expects structures to emerge in the future. Alternatively, Bob may find structures emerging quickly but foresee a limit to the extent to which this will continue. Here, our model predicts that he will be engaged but his interaction is unlikely to be sustained—we might imagine the 19 tone scale being written off as a 'gimmick'. Finally, suppose that Bob finds no structures emerging and also does not anticipate any ever emerging. In this case, the model predicts he will not be engaged and nor will their interaction be sustained.

This example illustrates the wide scope of IMSs that may be considered by the model, and how it may be used to form predictions based on a hypothetical interaction trajectory. It has also highlighted some of the limits of the model. The output of the system is considered as a uniform space leaving no scope for Bob to value some sounds more than others. Likewise, irrespective of prior knowledge, some types of gestural input may simply be more pleasing to Bob than others. At present these factors are not considered with the ES model. However, they remain important considerations for those designing IMSs and could be incorporated within future developments.

4.6 Exploratory behaviour

Having established in Section 4.4 the need to maintain both anticipated and realised increases in the participant's capacity for action, we now consider how differences in the participant's behaviour affect these two measures.

A spectrum between horizontal and vertical exploration was described in Section 4.1.2 as defined by two respective extremes of participants seeking out the largest number of 'playing techniques' and participants seeking to maximise what they could do with a minimal number of techniques. In terms of the ES model, horizontal exploration relates to seeking out as many shallow structures as possible through exposing inputs. Implications remain open with the possibility of deeper potential structures to be established over these inputs (and potentially a high encounter entropy too). Vertical exploration involves neglecting implications to establish deeper structures over that which has already been explored. As unexposed inputs have not been actively sought, we may argue that encounter entropy remains high. However, we may also argue that finding a great potential in one area of the input space creates expectation that similar potential is to be found over the remainder. In terms of the ES model this gives us

Prediction 14 *The emergence of deep structures over an area of the input space creates implications of similarly deep structures over the remainder of the input space.*

As implications anticipate their closure (Prediction 1), Prediction 14 implies that a participant who limits their exploration to a subset of the exposed input space establishes high anticipation of capacity gain.

Combining this with Prediction 6 gives us

Prediction 15 *A participant who limits their exploration to a small subset of the exposed input space with potential deep structures will tend to sustain interaction more than a participant who does not.*

What both horizontal and vertical approcahes have in common is that the participant is that the most easily attained increases in certainty are neglected. Always seeking to close implications and minimise uncertainty, as suggested by Prediction 2, does not lead to a continuing sense of potential from the system. This is in line with predictions from the CFD/CFI categorisation (Section 4.1.3) which suggested that curiosity driven by a feeling of being deprived specific knowledge (CFD) leads to a sense of closure rather than further curiosity. From this we may propose

Prediction 16 *For a fixed system, participants who seek to close implications without creating new implications will have a lower anticipation of capacity gain (i.e. a greater sense of the possibilities being exhausted).*

which, given Prediction 5 that anticipated capacity gain sustains interaction, we may tentatively suggest

Prediction 17 *For a fixed system, participants who prioritise minimising uncertainty will terminate their interaction more quickly than participants who prioritise maximising capacity.*

and as a result

Prediction 18 *For a fixed system, participants who prioritise minimising uncertainty will terminate their interaction with a lower capacity than participants who prioritise maximising capacity.*

Predictions 17 and 18 formalises the concern at the beginning of this chapter that those seeking to 'figure out' the system will do so and become bored without getting a sense of the potential of the system.

Why do these differences in participant behaviour arise? Litman (2010) argues that tolerance for uncertainty is a personality trait. Rather than instantly seeking answers, further questions are generated. However, with a better understanding of how it limits a participant from learning about the full potential the system offers them we are now in a position to do something about it.

At present, the predictions remain abstract. In the next chapter, we will derive a testable design principle from Prediction 15. However, before we do so, we will demonstrate how the ES model may be applied to understand the audience experience of the Serendiptichord—as well as a more subtle issue regarding anticipation that the model presently does not consider.

4.7 Example: The emerging structures of the Serendiptichord

We argued in Section 3.3.1 that spectating the Serendiptichord was a process of vicarious exploration. In this way, the Serendiptichord performance may be seen as a participant interacting with an IMS but with us controlling the behaviour of both so as to provide an 'idealised' exploration. Therefore, whilst the ES model is being developed to understand unscripted IMS performances, the Serendiptichord performance narrative (Section 3.2.5) is useful in illustrating how it characterises many of these aspects of an 'ideal' interactive music experience.

The Serendiptichord performance provides the audience with no prior knowledge. Before the instrument even makes an appearance, the dancer works to create a sense that there is something of interest within the box. This creates encounter entropy—we in the audience do not have any ideas of what is going to happen, but we anticipate that something of interest will happen. The use of the box allows encounter entropy to retained throughout the performance as it leaves open the possibilities that there are further physical components yet to be introduced which may then provide unexposed inputs.

The first inputs that are provided are simple movements to demonstrate that there is a mapping to perceive. It is necessary to be explicit here to due to the context. This creates implications that there are structures to emerge about how movement creates sound. These implications are realised as the dancer explores simple movements with the pod.

As the second pod is removed from the box, encounter entropy is maintained (or possibly increased) by the reinforced possibility that further parts of the instrument are to be introduced. The second pod is a mirror image of the first one creating an expectation that it will behave the same. However, as it is yet to do so this remains an implication. When the dancer shakes it in a similar manner as she did the first, nothing is heard and so this implication is not realised but remains open.

This surprise will have increased uncertainty in the perceived mapping creating a sense of tension—not only is the implication being sustained but the audience's expectation of a successful outcome (i.e. an increase in the dancer's capacity for action) has decreased. This tension is resolved and the implication realised when the two pods are used together and the greater structure emerges that the modules of the instrument interact in how they produce an output.

Throughout all of this, many of the finer details of the perceived mapping have remained unclear and the question of to what extent the dancer (and thus we) are really controlling the instrument remains only partially answered. Uncertainty over this issue is uncertainty over the entire exposed input space. Therefore it is a limitation of the dancer's capacity to act and anticipation of its answer is an implication. This (in my opinion) is realised when the trunk of the headpiece appears on the stage with its direct and reactive mapping.

We should point out that the ES model has been strongly informed by reflecting on the Serendiptichord so the above description is not really a validation but rather an illustration of how it may serve to understand the way in which musical interaction may construct a work in itself.

There is considerably more to the experience of spectating the Serendiptichord than is captured by the above analysis. We are not proposing that the model is a means to evaluate the success of a work. Rather, as we will demonstrate in the next chapter, we intend it to provide insight into one way in which a system may be designed to create a captivating experience.

4.8 The difficulty with interactivity

Thus far we have identified the need of a system to ensure the participants are expecting and attaining further gains in their capacity for action. However, the description of the Serendiptichord narrative has demonstrated a more subtle issue relating to anticipation—that of delayed gratification.

The participants of a non-expert IMS are presumably not trained composers. Pearce and Wiggins (2002) present a model of composing that involves creating a conceptual model of a listener's perception around which appropriate degree of expectedness, ambiguity and surprise may be crafted. To do so, the composition task is transformed from being ill-defined into being well-defined by establishing specific constraints within which to work (Pearce and Wiggins 2002). As composers become more experienced they are able to pay more consideration to the greater structure of their work, managing constraints at both the local and global level (Colley et al. 1992). Establishing a constraint across a work provides a redundancy within the work which is a potential structure. With interactive music, the 'composer' *is* the listener. We cannot particularly imagine them able to manage their own expectations and plan themselves surprises.[2]

Recall the argument from Huron's ITPRA model (2006, see also Section 2.4.3) that sustained expectation heightens arousal, amplifying the emotional response of its conclusion. We may therefore propose

Prediction 19 *The longer the anticipation of an equivalent increase in capacity, the more engaging it will be when it appears.*

This suggests that it is not enough to ensure the participant retains an anticipation of capacity gain but that specific implications are sustained over time.

Sustaining anticipation is key to establishing patterns of tension and release within music. Whether or not the participant is aiming to minimise uncertainty or maximise capacity, the kind of delayed gratification that lies at the root establishing anticipation is unlikely to play a part. Even if the participant did explore until they had a maximal mental model of the system, performance is quite different from demonstration. We could have *demonstrated* the Serendiptichord in less than a minute. Instead we drew it out, raising questions and delaying their answers and thus heightening the arousal—i.e. excitement—of their arrival. We may thus propose

Prediction 20 *A participant will be more engaged by a system if implications are sustained, i.e. the closure of implications are delayed.*

With an interactive system, the user holds both the questions and the means to answer them. There is little reason to expect them to delay their

[2]We are conflating the act of composing with that of improvising here. However, improvisation is as much a skill as composing requiring an intimate familiarity with both instrument and the style of music we are performing (Johnson-Laird 1988; Nachmanovitch 1990).

discovery any longer than necessary. Thus, as well as providing steadily paced emergence of structures, an IMS needs to establish establish sustained anticipation, i.e. implications of structures regarding specific inputs but without an immediately available answer. This provides us with a more readily applicable consequence of Prediction 20:

Prediction 21 *A participant will be more engaged by a system if they are made aware of inputs that will only be performable in the future.*

In addition to Prediction 15, this will form a part of the design principle we test in the next chapter.

4.9 Discussion

The Emerging Structures (ES) model describes the process taken by an individual exploring an IMS for the first time, focusing on aspects of an experience that arise through an increasing capacity to act. As mentioned following the example of Section 4.5, there most certainly will be other factors and processes involved in making a captivating interactive music experience. However, building on theories of musical appreciation and exploratory play, we have argued in this chapter that this process of forming an understanding of how a system works and how a participant may act through that system is a fundamental determinant of the participant's enjoyment. The predictions of Section 4.4 form a necessary precondition rather than a complete and sufficient formula.

The requirements identified by the model remain general. As such we have not specified any objective requirements of which IMSs it may be applicable to beyond the general definition provided in Section 2.2.3. One might question whether it is too general; for example, would it identify a fundamental difference between the two different experiences of interacting with an IMS and that of participating in an audience of a (not very good) human band? Whilst the ES model does not analyse what distinguishes these two experiences, it does not ignore the fact that such crucial distinctions will exist. The learning processes described within the model are defined with respect to both the prior knowledge of the participant and their cognitive abilities. Not only will an individual approaching either of these two experiences establish wildly different expectations about how the ‘system’ will behave; they will have differing mental faculties to draw upon in trying to understand them.

As a result, the analytical and predictive strength of the model is reliant upon the extent to which prior expectations and cognitive ability may be estimated and generalised across the participants of interest. In the sound installation example of Section 4.3 and the Serendiptichord example of Section 4.7 we were able to do so by considering IMSs that may be assumed to be completely unfamiliar to a typical participant. In the keyboard example of Section 4.5, we considered an IMS likely to be familiar to most participants but proposed a participant with specific (and unlikely) prior expectations. The processes described by the model remain important in scenarios where prior expectations or cognitive ability cannot be reasonably assumed, however forming specific predictions in these scenarios is unlikely to be practical.

We have been able to reason about such a broad range of IMSs by considering a participant’s understanding of the system in terms of mapping. As we saw in Section 2.2.7, concerns have been raised that analysing IMSs entirely in terms of mapping may be too reductionist. However, the perceived mapping that we are considering is not simply a pairing of input signals to output signals, but a probability distribution predicting the output of a system in response to the inputs a participant is aware of (see Section 4.2.1). This distribution, as well as what may be considered an input or an output, is defined by the participant’s mental model.

Whilst the ES model does not specifically model complex musical thoughts, arising from either participant or IMS, it places no limitations on the representations that may form the participant’s mental model or their complexity. Rather, it rests on the assumption that these representations exist, that they allow the participant to reduce their uncertainty over the output of the system given the inputs that they know, and that they develop over the course of an interaction based on the behaviour of the system as perceived by the participant in conjunction with their existing knowledge. As the perceived behaviour may be considered simply as concurrent sequences of inputs and outputs, the perceived mapping—a probability distribution over a space of outputs conditioned over a space of inputs—is the common medium through which a mental model is tested and improved. For this reason it remains the most suitable terms in which to define the model.

4.9.1 Quantifying Emerging Structures

Whilst defining the ES model in terms of subjective attributes is in line with the experience-oriented approach described in Section 1.3, quantifying the predictive power of the model is made more difficult as a result. However, the model is strongly rooted in ideas from information theory and its application as discussed in Section 2.4.3. This provides the possibility of creating a computational approximation of the model. For example, we could model the exploratory behaviour of a participant using machine learning tools from which we may estimate uncertainty, encounter entropy, implications (expected decreases in uncertainty), capacity for action (channel capacity). In this thesis, such an approach is left as future work (see Section 10.1.4).

4.10 Conclusion

In this section I have presented a new model of how a participant explores an Interactive Music System (IMS) for the first time: the Emerging Structures (ES) model. The ES model provides a common language that unites theories of exploratory behaviour in interactive art (Section 4.1) and those of musical perception (Section 2.4.3) to understand better the experience of interactive music. This has allowed a number of predictions to be made about how the behaviour of participant and system affect the participant's experience. We will be drawing on a selection of these as we progress in this thesis.

In the expectation-based models of Section 2.4.3, listening is motivated by our perceived potential to increase our ability to predict our environment. Here, exploration is motivated by our perceived potential to increase our ability to predictably *manipulate* our environment. With interactive music, we are *doing* as well as listening. We have argued that as well as involving unconscious learning, participating with an IMS is a more conscious process than non-interactive. In contrast with the aforementioned models, receiving continuous information from an IMS requires conscious involvement and decision making from the participant. Therefore we have placed a greater emphasis on understanding what *motivates* us to continue to be active in exploring such a system. In line with this, we have identified the perceived potential to improve our capacity to act, either through exposing new inputs or increasing our certainty over those that have been exposed. However, whilst interacting with a system may be a more conscious activity than simply listening to music, unconscious expectations remain important. We saw this in Chapter 2 with the prevalence of embodied knowledge (Section 2.2.6) and perhaps more dramatically with an increase in task performance with more complex mappings (Section 2.2.7). What was more complex to reason about was not necessarily more complex to control.

Placing the participant in control has raised unique difficulties not found in non-interactive media. As well as the potential for the participant not to discover aspects of the system, the composer no longer has the means to delay this discovery. This limits the potential to sustain specific anticipations and their potential to increase the pleasure of a resolution through focusing attention and heightening arousal (see Section 2.4.3).

The model presented here has allowed us to form the prediction that an interactive music experience will be made captivating through continually increasing the participant's perceived capacity for action and maintaining their expectation that further increases will emerge. Whilst we suggest that this is a difficult prediction to test directly, we have proposed that a system may be made more captivating if it encourages more *vertical* exploration—that is, if it leads to participants focusing on a subset of the input space in a way that maintains implications over the remainder of the exposed inputs. In the next chapter, we will derive a practical and general-purpose design principle from this and put it to the test.

Chapter 5

Emerging structures in practice

In this chapter I apply the abstract predictions of Chapter 4 to derive a simple design in which subsets of the input space are initially unperformable but gradually 'unlock' over the course of the participant's interaction. I implement this within a pair of simple Interactive Music Systems (IMSs) that provides control over a generative rhythmic loop system and evaluate it in a controlled user study. No significant effects are found. Through an analysis of the interaction logs and an informal analysis of additional qualitative data that was collected I argue that the lack of significance was down to different expectations among participants of what is an appropriate design for an IMS. In particular, a number of participants suggested that the manner in which constraints were implemented appeared 'unjustified'. This leads us to conclude that better tools are needed in order to reason about how participant expectations interact with how an IMS is perceived.

In the previous chapter I introduced an abstract model of how a participant explores a system, what engages them and what motivates them to continue. The model gave rise to a number of predictions. In this chapter, I derive from a number of these predictions a general design principle which we call *incremental constraint unlocking.*

5.1 Incremental constraint unlocking

We will be working primarily to apply the following two predictions from Chapter 4.

Prediction 15. *A participant who limits their exploration to a small subset of the exposed input space with potential deep structures will tend to sustain interaction more than a participant who does not.*

Prediction 20. *A participant will be more engaged by a system if they are made aware of inputs that will only be performable in the future.*

Recall also that we defined a *captivating* interaction as one that was both sustaining and engaging (Section 4.4.4).

These two predictions are complementary and suggest that we will increase the chances of creating a captivating experience if we

1. Constrain the performable inputs to a subset,
2. Create implications of the remainder of defined inputs, and
3. Make the participant expectant that they will become available in the future.

In effect the proposal is that by constraining the inputs that the participants may access they will explore with more focus, which will increase their perception of the system's potential. This will establish anticipation of realising this potential on the remainder of the input space when it becomes available, which will both enhance their appreciation of this functionality when it arrives and keep them engaged in the meantime.

Note, however, that these inputs do have to become available at some point. Not only is the participant unlikely to sustain their anticipation if the system never delivers the functionality they are expecting, but there is also a limit the potential depth that they are likely to perceive over the input space that is available.

Therefore, the approach we shall take is to construct a large system mapping but constrain the inputs the participant may perform to a small subset early on. Gradually through the course of their interaction, further inputs will become performable and so the size of this subset will grow.

Limiting the user to a enlarging subset of our interaction space is equivalent to restricting them from a diminishing subset. We will refer to the points within the restricted set as *constraints*, and the event where a point is removed from the set of constraints as *unlocking*. Through a process of incremental unlocking new inputs become available and hence potential structures.

However, the above points 2 and 3 require the locked inputs to be exposed and expectation of their unlocking established. Rather than hiding any feedback relating to unperformable inputs from the interface, their functionality and unavailability needs to be communicated.

5.1.1 Relationship with videogames

Incremental constraint unlocking is similar to the notion of levels in videogames. We reviewed music games in Section 2.2.5 but it is worth highlighting here how the vision of interactive music described over the previous two chapters compares with that of a game.

Games are a consumed experience. Although a player may take pride in their own mastery of a game, at no point is the player expected to take responsibility and 'use the game' as a means to help them have fun. The game is simply the source of the fun. In contrast, whilst a musical instrument may *facilitate* musical expression, it is not supposed to be its origin (Jordà 2004). Interactive music is also a consumed experience. Although we will likely wish to provide our user with a degree of creative freedom, responsibility for the quality of experience ultimately lies with the system.

We have observed that unrestrained exploration can make for a less enjoyable interactive music experience. Similar observations have been made of videogames (e.g. Clanton 1998; Rollings and Adams 2003). As with interactive music, efficiency is not the aim of the game designer. There is a need to establish anticipation: *sustained* expectations of both surprises and affirmations. However, a fundamental difference arises between games and interactive music with regards as to how this is done.

In games implications are sustained through setting challenges that the player must overcome through application of their mental model. Progress is determined by the player's performance allowing the emergence of structures to be paced according to their abilities. Having clear goals and reliable objective indicators of their accomplishment is an important criteria of good game design (Clanton 1998; Murphy et al. 2011). Even when the player does not know what they must do to advance within a game, there is still a clear challenge created by the implicit goal—find out what to do!

But with interactive music, goals are created by the participant rather than the designer. Indeed, we saw in the previous chapter that even interpreting there to be a goal of 'working out' the system may detract from a participant's experience. Our goals are often implicit and subjective: create a desirable musical outcome, explore the system, have an immersive experience. To some extent we may devise objective indicators of some of these (e.g. the proportion of potential inputs that have been received) but we then risk the possibility of establishing objectively-defined goals of 'overcoming' the system and our participant adapting their behaviour to fit within its (inevitably) imperfect indicators.

5.1.2 Timing

The question of *when* we unlock the constraints remains unanswered. Prediction 10 proposed that the participant will be more engaged if structures are continually emerging. This suggests that the optimal moment to unlock the constraints are at the point when the rate at which uncertainty is decreasing drops below a threshold. However, structures exist in the participant's mental model so we are unable to determine when this is. We could attempt to estimate it through defining some objective indicator over their behaviour, but as we saw in Section 5.1.1, this then runs the risk of establishing an objective goal that the participant seeks to satisfy as opposed to maximising their capacity for action.

For these reasons, we will leave the question open and, for now at least, opt for a simple timer-based solution. In Chapter 7, I will propose a more refined approach.

5.1.3 Summary

Incremental constraint unlocking

The input space is partitioned into two sets: 'constrained' and 'unconstrained'. Constrained inputs are unperformable. At regular moments through the interaction, constrained inputs are unlocked*—i.e. removed from the constrained set into the unconstrained set.*

The interface is constructed to create implications of the output associated with constrained inputs, as well as establishing expectation in the participant that the constrained inputs will become unlocked.

5.2 Evaluation criteria

There are a number of aspects of a participant's experience that may be affected by the design principle. These were chosen to be those most frequently sought within the NIME community. They are as follows.

Transparency (T) is how intuitive the connection between the input and output of the system feels to the participant (Section 2.2.10). In terms of emerging structures (Section 4.2) we may see this as the average degree of certainty within the participant's perceived mapping over the course of the interaction.

Learnability (L) is how quickly the user feels they are learning to use the system (Section 2.2.12). We may see this as an estimation of the average rate at which structures emerged.

Diversity (D) is the perceived range of potential musical outputs of the system (Section 2.2.11), which we may associate with their belief of the potential capacity of the system.

Autonomy (A) is the extent to which the participant feels that *they* are responsible for the output of the system.

Quality of output (Q) is the participant's opinion of how good the music produced during their use of the system is.

Fun (F) is how much the participant enjoys using the system.

Engagement (E) is how focused and interested the participant is while using the system.

Therefore, drawing on our proposals of Chapter 4, we predict the following effects.

- Transparency, Engagement are the key criteria in which we would expect to see an increase.
- Learnability we would also expect to see an increase in given the argument that vertical exploration promotes deeper exploration (see Section 4.6).
- Fun and Quality of output we might tentatively hope for an increase, drawing on our observations of investigative and diversive exploration (Section 4.1.1) and the problem of anticipation within interactive systems (Section 4.8) respectively.
- Diversity and Autonomy are two topics which we feel may be affected by our design principle. Although incremental constraint unlocking quite clearly reduces the *objective* diversity of outputs and the control granted participants, we do not know whether the participant will perceive as such.

The above measurements are clearly not independent. For example, we would expect transparency and learnability to be related. We will investigate correlations between our hypotheses in Section 5.5.1.

5.3 Method

5.3.1 Methodological issues

Empirical evaluation of such a design principle presents a particular challenge. When measuring factors such as transparency, learnability and engagement, we would expect a user's experience to vary with repeated use of the software leading to a unacceptably large ordering effect. This would suggest a study design of non-repeated measures is necessary to prevent an ordering effect between the two conditions.[1] However, due to individual differences among participants' taste and technical background a large amount variation is to be expected among how participants respond to musical stimulus as well as software requiring a creative input. This high variance means that within a non-repeated measures design, a substantial number of participants would be needed to achieve significance. For example, Gonzales et al. (2009) found a sound installation was significantly more enjoyable with an interactive component than without ($p < 0.01$) in a non-repeated measures study which recruited 142 participants. As the difference between our two conditions is not quite as dramatic as adding or removing interaction, we may anticipate that in our case effects would be less consistent, requiring more participants than this. Such a large number of participants was not practical at this stage in the research.

With a repeated measures design, variance within the sample is less of a problem as we are only comparing each individual's change in response. Therefore, this design was chosen. However, in order to address the above

[1] Under a non-repeated measures design, each participant is recorded under a single condition with data compared *between subjects*. In a repeated measures design, data is collected from a participant under all conditions and compared *within subjects*.

concerns regarding ordering, the following specialised study design was created in collaboration with Prof. Rosemary Bailey, a statistician with expertise in designing experiments that compare qualitative treatments on difficult to measure effects (Bailey 2008).

5.3.2 Study design

Two interactive music systems, α and β (described below), were developed to be as similar as possible in functionality whilst being different enough in design to minimise the effect that using one would have on a subsequent experience of the other. Participants would then experience one system with incremental constraint unlocking (condition A), and one without (condition B). In order to be able to measure and mitigate any ordering effect, as well as any bias caused by differences between the systems, each combination and ordering of conditions and systems was used. That is, system α under condition A followed by system β under condition B; αB followed by βA; βA followed by αB; and βB followed by αA. For brevity, we will refer to a system with incremental constraint unlocking as a *constrained* system.

Any effects as described by the hypotheses were measured by means of a questionnaire, described below.

Prior to using each system, participants were given a brief overview demonstrating which interactions were available within the interface (i.e. which interface elements were buttons and what could be dragged). The effects of these actions were not explained beyond the fact that nothing would be heard until an 'on' button was pressed and the functionality of the 'snapshot' feature (explained below). In order to limit the possibility that participants thought that the constrained system was functioning incorrectly, where a system was constrained this fact was explained during the overview. Following both conditions, participants were given the questionnaire to fill out, followed by a short structured interview.

Further data was collected by the system which logged all interactions including mouse movements and button presses and releases.

5.3.3 Participants

Twenty-four participants were recruited for the study via a department-wide email. They were paid £7.50 for approximately an hour's involvement. Most were students within the university's School of Electronic Engineering and Computer Science.

A pilot study indicated that the musical experience of participants may affect the efficacy of constraint unlocking so this was also included as a controlled factor. Participants were recruited via email and asked to answer a number of multiple choice questions regarding their musical experience, which was used to provide each with a musical score (see Appendix B.1). Potential participants were then divided into two near equally sized groups of *musically experienced* and *musically inexperienced.* An equal number was selected from each, who were then randomly allocated into four equally sized groups corresponding to the four system/condition combinations described above.

5.3.4 Questionnaire

To avoid the inconsistency that may arise with self-reported quantitative data (e.g. the Likert scale (Jamieson 2004)), participants were asked only to provide comparisons between the sessions. The questionnaire presented a series of statements, for each of which the participant was asked to select whether the statement most applied to the first session or second session as well as provide a brief explanation. For each hypothesis, four statements were included—two worded positively and two negatively giving 28 in total. Each statement could either support or oppose its respective hypothesis (or be left blank) giving an integer score between -4 and $+4$ which was then scaled to the range $[-1, 1]$. Although only providing nine possible values for each hypothesis, this method provides interval data allowing parametric significance tests to be used.

A list of the statements alongside their respective hypotheses is provided in Appendix B.2.

5.3.5 Interviews

The short structured interview was intended to allow participants to respond more openly to the answers they provided in the questionnaire, as well as to reveal the purpose of the study and probe their opinions of incremental constraint unlocking more specifically.

The following questions were asked in the interview:

- Which session did you prefer?
- Why was that?
- In which session did you put more effort into creating a musical structure?
- In the first/second session the interface was constrained initially. How did this affect your experience?
- Did you feel a sense of anticipation for the locked features?
- Did the unlocking give your piece more structure?

5.3.6 Two interactive music systems

In order to achieve the aim of creating systems that had similar functionality whilst being different to use, two interfaces were developed that both controlled a common underlying generative algorithm: the *Manhattan Rhythm Machine* (MRM), which was then used to control two different sets of MIDI instruments. The MRM is a deterministic and continuous mapping from four-dimensional space to a monophonic rhythm with varying note velocities. These four dimensions were named based on their effect on the output as NUMBER OF NOTES, EDGINESS, COMPRESSION and VOLUME. A further discrete parameter controlled the length of these sequences (either 4, 8, 16 or 32 semiquavers).

The algorithm was used to control five instruments: hi-hat, snare hit, snare rim-shot, kick drum and a bass synth following a simple sequence of pitch values. At the start of each session, each instrument was initialised to a distinct state that was standardised across all sessions. Each instrument could be switched on or off and the state of the entire system could be saved or loaded into six 'snapshot' slots.

The same set of constraints were used in both interfaces. Denoting the continuous parameters of each instrument as $(x_1, x_2, x_3, x_4) \in [0, 1]^4$, then these were as follows.

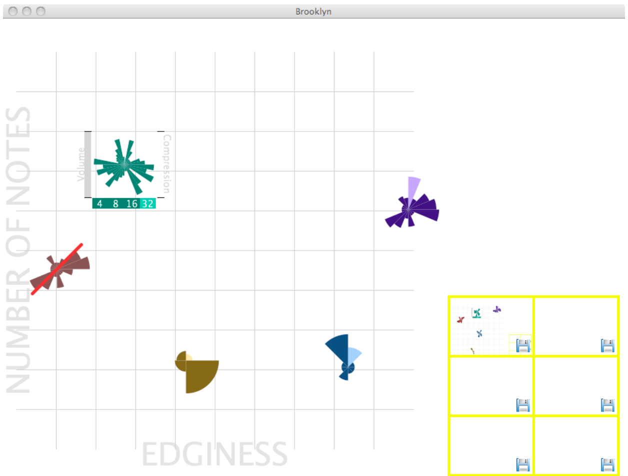
- The bass synth is disabled and for all instruments x_1 and x_3 are locked to their original (instrument-specific) value until bar 33.
- The snare drum is locked and for all instruments sequence lengths are fixed to their original value, x_1 is restricted to values within 0.3 of its original value until bar 65.
- For all instruments the value of x_1 is restricted to values within 0.7 of its original value until bar 97.

When incremental constraint unlocking were in operation, a message was shown for eight bars prior to an unlocking informing the participant of what was going to be unlocked, and eight bars following an unlocking informing them of what had been unlocked. The interfaces were both designed to make it clear when items were locked showing either the word 'LOCKED' or a picture of a padlock over the relevant feature. Figure 5.1 shows both interfaces used with and without constraint unlocking.

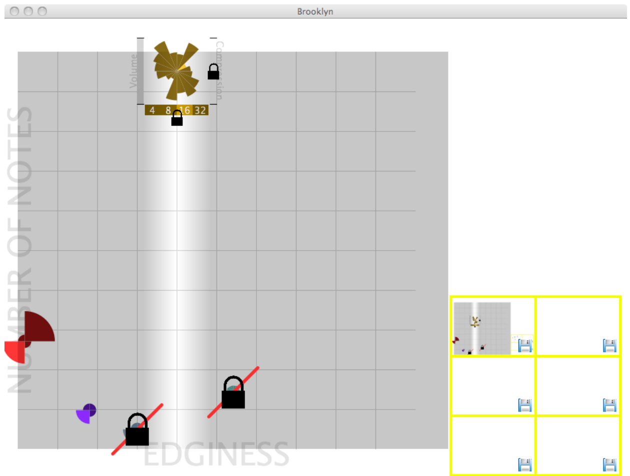
Participants were given seven minutes (208 bars) to use each interface. For the final 30 seconds of each session a message was displayed asking them to wrap up their piece of music.

5.3.7 Log file analysis

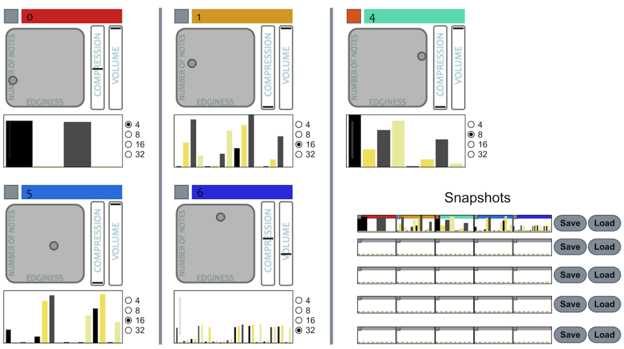
The log file for each session was analysed to determine the rate at which it was explored. Two quantitative measures were created.



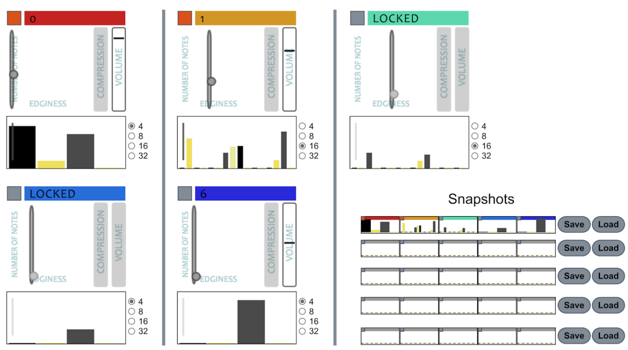
(a) Interface α without constraint unlocking.



(b) Interface α with constraint unlocking.



(c) Interface β without constraint unlocking.



(d) Interface β with constraint unlocking.

Figure 5.1: The two interfaces used within the study shown under both experimental conditions.

Duration until maximal feature set used

For the first we considered the point at which a maximal set of 'features' were used by the participant. The following were identified as features.

1. Turning on an instrument (one feature for each instrument),
2. Modifying the (individually controlled) x_3 or x_4 parameter of any instrument by 0.1 or greater[2] (two features),
3. Changing the sequence length of any instrument (one feature),
4. Moving any instrument within one of the four quadrants of the two-dimensional space defined by x_1 and x_2 (as these parameters are always controlled by a two-dimensional slider) (one feature for each quadrant),
5. Saving and loading a snapshot (two features).

Not all features as defined above were used in every session. Therefore we considered the length of time until a maximal set of features was used, i.e. the point when all features that were going to be used, were used.

Time spent in novel states

Considering the state space of the system as a multi-dimensional vector space, we may then consider *how* participants explored it by analysing their path through this vector space.

With five instruments, each with four continuous parameters and two discrete parameters, the space of the system may be represented by a 30-dimensional vector, with 20 components varying continuously in $[0, 1]$, five components (on/off switches) varying in $\{0, 1\}$ and five components (sequence lengths) varying in $\{4, 8, 16, 32\}$. With continuous parameters, it is unlikely

[2]Recall that our continuous parameters are all defined over the range $[0, 1]$.

that many states will be occur more than once[3] but by quantising the continuous parameters we may gauge how long participants are spending within a particular area of the state space. Lengthier stays would suggest a deeper and more focused exploration.

Each continuous parameter was quantised into four bins giving a space with a total of 2^{55} discrete states. State vectors were then recorded at moments where the (underlying continuous) state changed alongside the time spent in that state.

We define a quantised state as being *novel* if it has not been visited before. From this we may consider what proportion of time is spent in novel states, and what proportion in states that had been previously visited (after quantisation). We refer to this measure as the *time spent in novel states.*

5.4 Hypotheses

Seven hypotheses corresponding to the above evaluation criteria (Section 5.2) were tested in this study: *Participants using a system with incremental constraint unlocking implemented will perceive*

(T) an increase in transparency,

(L) an increase in learnability,

(D) a change in diversity,

(A) a change in autonomy,

(Q) an increase in quality of output,

(F) an increase in fun,

(E) an increase in engagement.

[3]Technically in a continuous space we would expect the probability of the same state occurring more than once to be zero. However, as our input was determined by the pixel position of the mouse cursor on the screen, the parameters are only a discrete approximation of continuous state and hence repetition is possible.

Each statement the participant responds to in the questionnaire ranks the sessions. We have an equal number of positive and negative statements for each hypothesis. Therefore under the null hypothesis for each of the above alternative hypotheses we would expect a score of zero.

5.5 Results

5.5.1 Quantitative data

Overview

All measurements taken as described in Section 5.3.4 were scaled to the range $[-1, 1]$. Table 5.1 shows the unpartitioned results recorded.

Table 5.1: An overview of the questionnaire data without any partitioning of participants. Note that as each datapoint could only take nine possible values, some numbers are identical.

Hypothesis	Min	Max	Mean	Standard deviation
Transparency	-1.	1.	-0.042	0.77
Learnability	-1.	0.75	-0.010	0.48
Diversity	-1.	1.	-0.10	0.57
Autonomy	-1.	1.	-0.10	0.67
Quality	-1.	1.	-0.10	0.80
Fun	-1.	1.	-0.042	0.71
Engagement	-1.	1.	-0.042	0.62

The mean of each measurement is shown in Figure 5.2 with a 95 per cent confidence interval calculated using a one-sampled t-test. This simple overview gives a sense of the variance of the data but does not take into consideration the other controlled factors as the ANOVA does below.

Five participants (all in the non-experienced group) recorded a negative score for the constrained system on all seven measurements. One (also in the non-experienced group) recorded a positive score on all seven measurements.

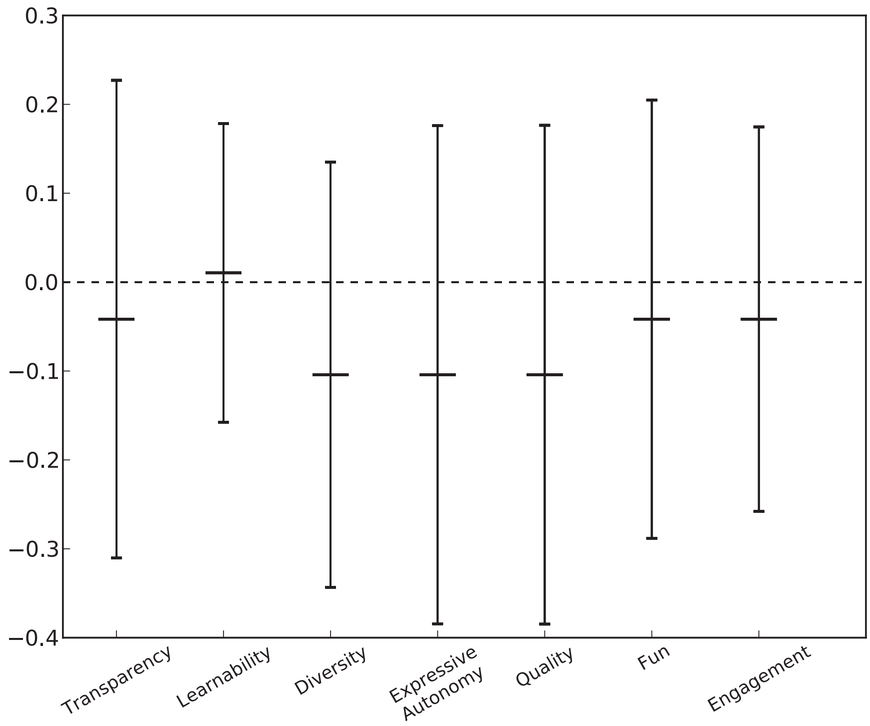


Figure 5.2: The mean value recorded in the questionnaires for each hypothesis, shown with a 95 per cent confidence interval. The value of zero is expected by the corresponding null hypotheses.

ANOVA of questionnaire data

The questionnaire data collected from participants was analysed using a first-order ANOVA,[4] which considered the effect of the dependent variable

[4]ANalysis Of VAriance (ANOVA) is a statistical test that determines the likelihood of two samples being drawn from the same population by considering how much their variances differ from each other.

(whether or not the system was **constrained**) and the other controlled factors (the **order** in which the dependent variable was presented to the participant, which **interface** was used in the constrained condition and whether or not the participant was in the musically **experienced** group). No significant results were found at $p < 0.05$.

Following this, a second order ANOVA was used, which also considered interactions between the controlled factors (e.g. whether experienced participants were more likely to show an ordering effect). No significant results were found at $p < 0.05$. Table 5.2 shows the F-ratios and p values found in this analysis.

Typically, we may interpret an F-ratio above 1 as indicating that with an increased number of participants we may well find a significant result. As can be seen, for our main experimental condition, **constrained**, none of the F-ratios are above 1 suggesting that the number of participants is unlikely to have been the cause for this lack of significance.

Relationships between the hypotheses

As observed in Section 5.2, there are likely to be dependencies between the hypotheses. Identifying these provide a useful means to consider the trade-offs and mutual dependence of different design decisions that may be taken when building such a system. A correlation matrix showing the Pearson correlation coefficient between each pair of hypotheses is shown in Figure 5.3.

Transparency and Learnability are highly correlated as would be expected. Most of the remaining hypotheses show a moderate correlation, with the exception of Diversity with Transparency and Learnability which is to be expected as the greater the range of outputs, the more complex a system is to learn.

It is of interest that the relationship between Fun and Diversity is considerably stronger than that between Fun and Learnability, indicating that the anticipated capacity of the system perceived by a participant may be more important to their enjoyment than the rate they felt able to attain

Table 5.2: Results of a second order ANOVA over the questionnaire data.

Hypothesis	Factor	F-ratio	p	Hypothesis	Factor	F-ratio	p
	constrained	0.067	0.80		constrained	0.0095	0.92
	interface	1.7	0.21		interface	1.6	0.22
	order	0.067	0.80		order	0.24	0.63
Transparency	experience	2.4	0.14	Learnability	experience	0.47	0.50
	interface & order	0.60	0.45		interface & order	0.77	0.39
	order & experience	0.067	0.80		order & experience	0.0095	0.92
	interface & experience	1.1	0.31		interface & experience	0.24	0.63
	constrained	0.87	0.36		constrained	0.50	0.49
	interface	2.2	0.15		interface	0.079	0.78
	order	0.31	0.58		order	0.18	0.68
Diversity	experience	0.31	0.58	Autonomy	experience	0.32	0.58
	interface & order	3.5	0.79		interface & order	0.32	0.58
	order & experience	0.31	0.58		order & experience	0.71	0.41
	interface & experience	2.2	0.15		interface & experience	1.6	0.22
	constrained	0.33	0.58		constrained	0.085	0.77
	interface	0.64	0.44		interface	0.34	0.57
	order	1.1	0.32		order	2.1	0.16
Quality	experience	0.33	0.58	Fun	experience	0.085	0.77
	interface & order	0.18	0.74		interface & order	0.34	0.57
	order & experience	0.18	0.74		order & experience	0.34	0.57
	interface & experience	0.18	0.74		interface & experience	4.2	0.057
	constrained	0.0095	0.92				
	interface	1.6	0.22				
	order	0.24	0.63				
Engagement	experience	0.47	0.50				
	interface & order	0.77	0.39				
	order & experience	0.0095	0.92				
	interface & experience	0.24	0.63				

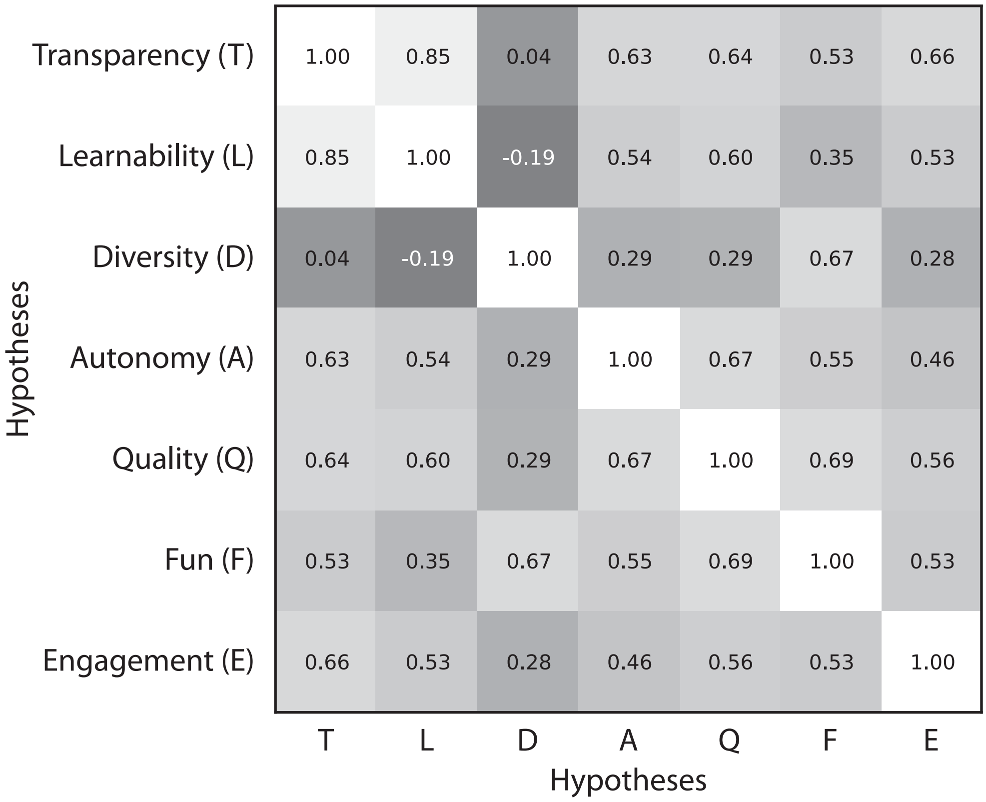


Figure 5.3: The correlation coefficient between each pair of hypotheses.

it. The opposite trend is shown when comparing Engagement to Diversity and Learnability. This offers mild support for our prediction that it is realised rather than anticipated capacity gain that makes interaction engaging (Section 4.4.3).

Principal Component Analysis

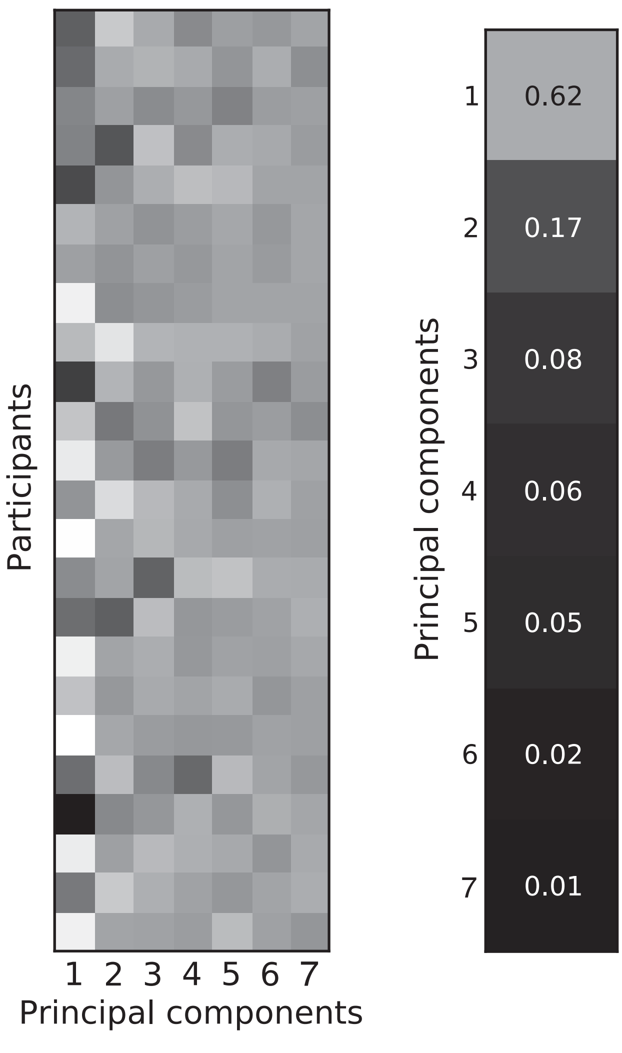


Figure 5.4: The questionnaire data projected onto its principal components and the proportion of the variance explained by each principal component.

As we saw in Section 5.5.1, the data collected had a large degree of variance. Whilst looking at correlations between our hypotheses measurements is of interest to see how different factors relate, it is useful to get a feel for how they combine to create this variance. A principal component analysis (PCA) was performed to identify the likely number of independent dimensions within the questionnaire data. The result of this analysis is shown in Figure 5.4. Recall that the questionnaires measured how far an aspect of user experienced *changed* between the conditions. The principal components are best understood as aspects of the individual participants that affected their response to constraint unlocking.

With 89 per cent of the variance explained by the first two principal components, it is reasonable to assume that the underlying variable we are measuring is two-dimensional with the remainder of the variance caused by measurement error. Figure 5.5 shows how much of each hypothesis measurement makes up these first two principal components. From this, we see that the first component, representing the majority of the variance, appears to be a general 'goodness' score that affects each aspect of our evaluation criteria in a consistent way. Thus, we may interpret 62 per cent of our variance as originating from an unexplained trait among our participants that influence how well participants respond overall to a constrained system.

The second dimension indicates that a trait may determine the extent to which a participant's answers demonstrated a tradeoff between Transparency/Learnability and Diversity/Fun. Seventeen per cent of the variance is explained by how participants responded to this conflict but, interestingly, Autonomy, Quality and Engagement did not seem to be affected by it. A second order ANOVA on the projected data found that which interface was constrained had a significant effect ($p < 0.05$, $F = 4.46$, $df = 17, 1$) on how participants responded within this second component with those higher responses from those for whom β was constrained.

An F-Test found that the musical experience of a participant had a significant effect on how they responded within the first principal component

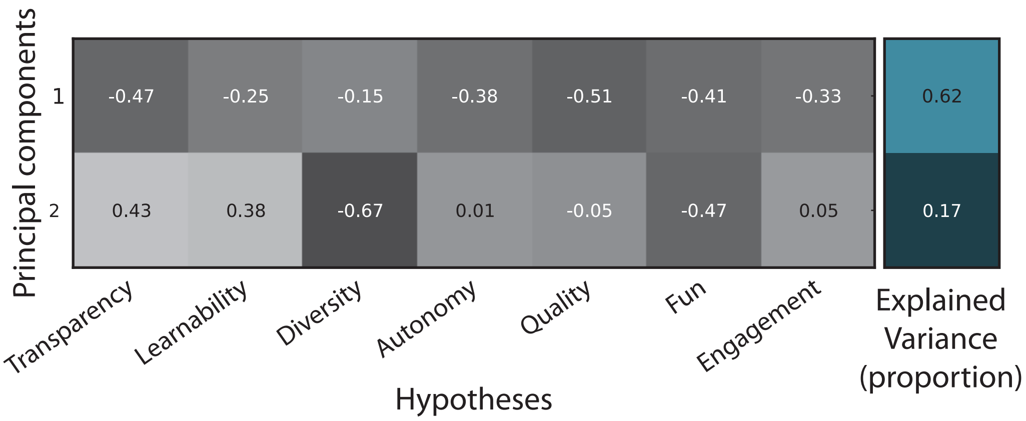


Figure 5.5: The transformation matrix mapping the questionnaire data to the first two principal components. Note that the overall sign of each row of the matrix is arbitrary.

($p < 0.05$, $F = 2.94$, $df = 11, 11$).

5.5.2 Qualitative data

Qualitative data was collected in this study both through the interview as well as space provided by each statement asking participants to explain their choice. Participants provided answers for 60 per cent of the statements.

This data was analysed informally in order to better understand the causes behind the quantitative results of the study.

Selected questionnaire explanations

We will select here a few explanations to explore further how a few individuals responded to constraint unlocking. We will consider the participants who responded in the extreme ends of the spectrum as well as a participant from the middle. Participants are referred to by the letter P and the number they were assigned in the study.

P24 who used β constrained followed by α unconstrained, responded the least positively to the constrained system, giving the unconstrained system a higher score on the evaluation criteria 89 per cent of the time. The most prominent theme in their explanations is the simplicity of second system. Regarding the first these included: 'was not getting the desired settings due to the complexity,' 'too many different windows made it complex.' Regarding the second: 'single structure, not too many complications,' 'it was easy to use and has more options.'

P40 (α unconstrained then β constrained) responded the most positively to the constrained system, giving it a higher score 96 per cent of the time. Their explanations mainly refer to the range of options available. Referring to the second session, these included 'more options regarding mixing were available and minute things could be changed,' it 'was according to my taste,' it had 'a good broad range.' Referring to the first session their comments included 'could not understand initially how everything is working.'

P23 (α unconstrained then β constrained) gave the constrained system a higher score 50 per cent of the time, which is very close to the median. In general, their explanations focus on differences between the interfaces with β's more conventional view aiding transparency ('more of a channel strip approach which I am used to') but being less engaging ('The look is familiar so less concentration was needed') whilst commenting on the novelty of α ('More interesting than B'). However, they did refer to constraint unlocking as having a positive impact on one engagement statement: 'Being forced to deal with only 2 instruments at a time helped me to focus.'

Following the above observation that the second principal component of our data seems to show a response to the Transparency/Diversity trade-off, it is of interest to see the explanations that participants at the extremes of this measure gave.

P04 (α constrained then β unconstrained) gave the lowest response on this component indicating that they perceived higher Diversity/Fun and lower Transparency/Learnability with the constrained system. They responded slightly in favour of constraints, rating that system more favourably 61 per cent of the time. Their explanations refer a number of times to the effect of the constraints on their exploration of the first system: 'having things unlock throughout made me explore each one more,' 'the nature of the first interface encourages playful exploration,' 'I didn't feel like I'd exhausted all the possibilities, 'fun to figure out.' That the second system was unconstrained did not seem to hinder its ease of use ('I probably had more control'), but did seem to decrease its perceived diversity: 'It wasn't long before I felt I was just going round in circles.' Although their explanations suggest a positive experience with the constrained system as it encouraged a deeper (and initially narrower) exploration, overall they reported it as slightly less involving as they were 'sort of waiting for things to become available.'

P15 (β constrained then α unconstrained), who responded positively to the constrained system 34 per cent of the time gave the highest response on this component. Again, most statements referred to the diversity of the systems: the second had 'a greater variety of sounds,' 'more scope for variation,' 'more surprises' and was 'more interesting.' Again, lower learnability and transparency was not a problem: they enjoyed themselves less in the first session because it was 'not as unpredictable' whereas in the second they reported feeling slightly more creative 'once I got the "hang" of it.'

Summary of interview responses

Of the 24 participants, 9 responded that they preferred the constrained session and 12 the non-constrained session. Similarly, 9 responded that the constraints had a positive effect and 10 that they had a negative effect. Of the 9 that were positive about the constraints, 4 mentioned that they added some kind of structural aspect to the piece. Of the 10 that reported a negative effect, 5 found them annoying because they prevented them from doing what they wanted and 4 found that they made the interface more confusing.

Of the participants who found the constraints annoying, the general consensus was that they were an 'imposition' that seemed like an artificial imposition by the system's designer. In contrast, one participant who appreciated the constraints mentioned that he was a regular videogame player so he was quite happy to work with these kinds of limitations.

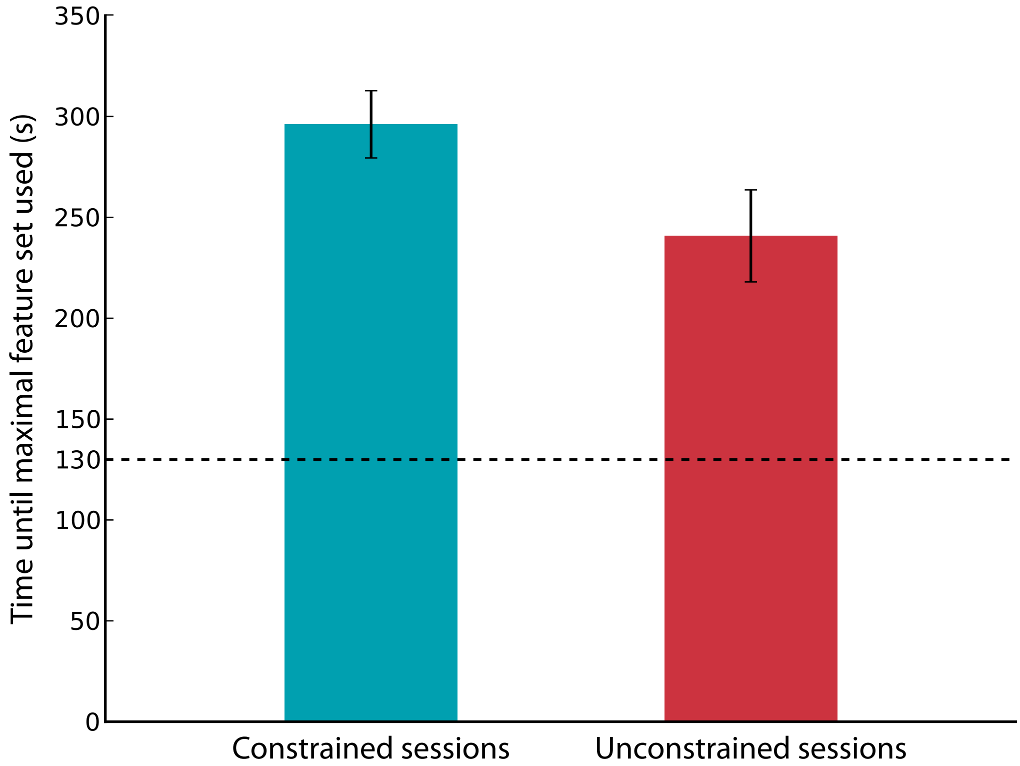


Figure 5.6: The mean length of time in seconds until a maximal feature set was used in sessions with constraint unlocking against sessions without constraint unlocking. In a session with constraint unlocking, all features as defined here would become available after 130 seconds. Error bars indicate the standard error.

5.5.3 Log file analysis

Duration until maximal feature set used

With exploration measured by the duration until a maximal feature set was used (see Section 5.3.7), we found a clear increase when constraint unlocking was being used, as shown in Figure 5.6.

Figure 5.7 shows the same measurement, but with the data partitioned by whether participants were in the experienced group or not. We find that the experienced participants showed a smaller increase between the sessions.

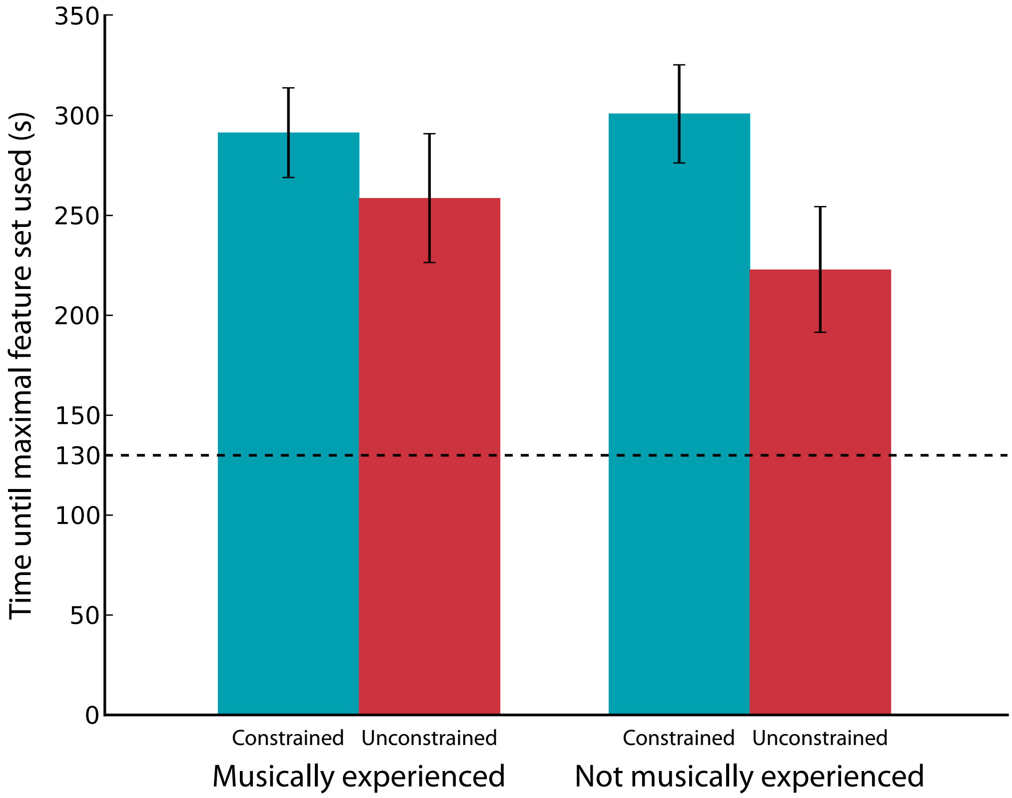


Figure 5.7: The mean length of time in seconds until a maximal feature set was used in sessions with constraint unlocking against sessions without constraint unlocking. Sessions have been partitioned by whether or not the participant was in the musically experienced group. Error bars indicate standard error.

This measure of exploration suggests that, for inexperienced participants at least, the implementation of constraint unlocking may have decreased the rate at which participants explored.

Time spent in novel states

Figure 5.8 shows the correlation between an increase in the length of time spent in previously unvisited states and the questionnaire data. The connection is generally moderately weak (and not statistically significant) except for a negative correlation with Diversity (D) ($r = -0.45$, $p < 0.05$). Participants who spent more time in novel states—i.e. explored broader—perceived there to be a narrower range of possibilities.

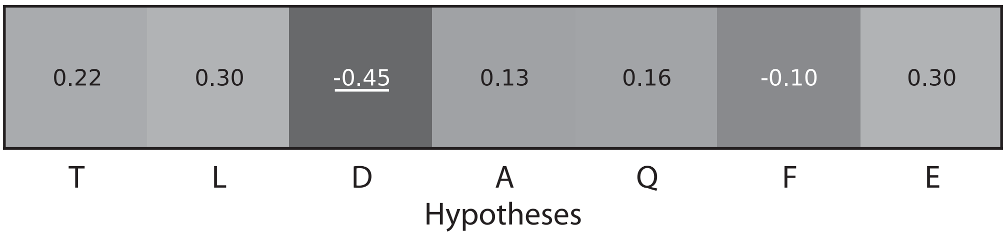


Figure 5.8: The correlation coefficient of the questionnaire data against an increase in the amount of time spent in a novel state (after quantisation). Underlined values are significant at $p < 0.05$.

5.6 Discussion

Whilst the study did not reject any of its null hypotheses, further analysis of both the log files and questionnaire data yielded a number of interesting findings.

The variance among the responses participants gave is high. This has been supported by informal analysis of the qualitative data which suggests that incremental constraint unlocking did have a noticeable effect for most participants. However, depending on the individual, this effect went in either direction preventing this from showing within the aggregated quantitative data.

Performing a principal component analysis allowed us to identify the most prominent ways in which these individual differences showed. We found a primary component which led to a universally positive or negative response to constraint unlocking. Interestingly, however, the second component explaining 17 per cent of the variance within the data suggests a trait that determines how a participant responds to a potential trade-off between *easy to understand and learn* and *diverse in potential outputs and fun* but independent of how engaged or in control they felt.

Analysis of the log data indicates that the design principle succeeded in slowing down exploration. This is not surprising given that our method of measuring exploratory rate is directly related to how we implemented the constraints. However, it is worth noting that features remained locked long enough to actually slow our participants. More interestingly, the effect was more pronounced among non-musicians, suggesting that those with musical experience tend to explore slower anyway.

We also observed that those who spent more time in novel states reported the diversity of the system as lower. At first this may seem somewhat counterintuitive, however recall that diversity is the perception of the *potential* range of the system. Therefore, we may conclude that the perception of the potential of the system depends more on what remains unexplored rather than what has already been tried. In terms of the Emerging Structures (ES) model, it is the anticipation of future gains in capacity rather than realised gains that determines the perception of the system's potential. This is in line with our argument that anticipation is key to avoiding the perception of 'exhausted possibilities' and sustaining interaction (Section 4.4.2).

One reason for the lack of significant effects observed may be that whilst this implementation of constraint unlocking did reduce the rate at which participants could provide inputs, it did not correspondingly reduce the complexity of the behaviour of the system. For example, the implementation of constraints did not hide any aspects of the interface. As a result, whilst the participants were only exploring a limited area of the input space, the potential structures that were available with which to construct a mapping may not have been reduced. The behaviour of the system may then have remained too complex for the participant to perceive sufficient structures within its behaviour to create any implications (i.e. anticipated reductions in uncertainty). In other words, they are overwhelmed, become frustrated by slow progress and do not anticipate this improving when the constraints unlock.

An alternative possibility is that the total system mapping was not complex enough. Recall that Prediction 15 (Prediction 4.6) required potential deep structures over the subset of the input space that the participant was exploring. It may be that the constraints did encourage more focused exploration but that this did not yield any benefits. In other words, they are underwhelmed.

However, whilst the unvarying complexity of the systems remain a possible confounding factors, the variance of the responses to constraint unlocking is suggestive of a factor among the participants themselves that determined their response. In particular, half of those who disliked the constraints reported them as 'annoying' and there was a sense that the designer was unjustified in imposing them upon them. This viewpoint is perhaps understandable but problematic if we wish encourage a particular type of interaction. What does it mean for an aspect of the design to be 'unjustified'? If the participant has no prior experience of a system, why does limiting certain inputs seem an imposition? Would it still have seemed the case if we had not marked the constrained aspects of the interface as 'LOCKED'?

Participant preconceptions of what a system 'should' be like were not considered within the ES model. Therefore, in the next chapter we will outline the second key theoretical component of this thesis, a framework of *Perceived Agency*. This will allow is to more rigorously analyse what it means to perceive constraint unlocking as an imposition from a third party. More crucially, it will allow us to identify how the expectations of the participant of *what the system is for* may affect their experience of interacting with it.

An empirical study such as this is designed to be robust against mistakenly reporting a positive result. Where statistical significance has not been found, we must rely on other means to determine whether the cause was a faulty premise, this specific implementation of that premise or a deeper methodological problem. We have explored the first of these two items here. However, there are also issues with methodology specific to studies involving creative or aesthetically-driven systems. These will be discussed in detail later on in Section 8.1

Chapter 6

Perceiving agency

In this chapter we address the question of why the previous implementation of constraint unlocking may have been perceived as 'unjustified' imposition by the participants. To answer this we consider what kind of performative freedom is necessary within a musical act and correspondingly whether skill needs to be demonstrated. I argue that, for Digital Musical Instruments at least, skill is a necessary component. To better understand the relationship between skill, freedom and how these are perceived, I present a new framework of Perceived Agency and use it to explore how different contexts of performative and participatory music require skill to be demonstrated in different ways. We conclude that although there are a diversity of norms of perceived agency, unexpected deviation from this norm may be interpreted as dishonest, inauthentic or—as in our case—an unjustified imposition.

In the previous chapter we found that many participants described the constrained interaction design as unjustified or an imposition on what they could do. But this was not the case with all the participants. Some were content to work within the constraints. In this chapter we will seek to develop a more rigorous understanding of what it means for an aspect of an Interactive Music System (IMS) to be unjustified and why this was only perceived to be the case by a number of the participants.

There are also a number of questions throughout this thesis that have remained unanswered. The IMSs of this thesis are for non-experts. However, in Chapter 3 we saw the possibility that demonstrating skill might be a necessary component of musical expression. If that is the case then will our participants inevitably not be able to express themselves with an IMS? If they can express themselves, do they need to be creative in doing so? Perhaps more importantly, do they need to *feel* creative? Do the answers to these questions depend on the particular ambitions of the IMS designer or are there general principles that may answer them? Does the participant need to be aware of the answers? It is through addressing these questions that we will gain a greater understanding of what a participant feels a system 'should' do and therefore why its behaviour may be perceived as an imposition rather than a feature or quirk.

Within the context of participatory music, questions such as these regarding the role of the user remain unanswered (Leong et al. 2011). Therefore to address them, we will once again examine Digital Musical Instrument (DMI) performance. DMIs offer a particular insight into this issue due to the extensive research done on attempting to make them as acceptable to musicians and audiences as acoustic instruments. It is, in fact, through considering the diversity of different roles a DMI can play that we identified in Chapter 3 that we may demonstrate the significance of perceiving skill.

Following a discussion of skill in DMIs, I will present a new framework that refines the notion of skill by considering different types—creative, technical and expressive, different ways in which skill may be demonstrated—prepared or live, and different individuals who may be skilful within a performance or interaction. Most importantly, however, it will allow us to understand more clearly how a performance or interaction context is perceived to grant *agency*—the freedom and responsibility to realise these skills, and what consequences this may have when a performance is thought to deviate from these norms.

With this framework of perceived agency, we will then be able to clarify what it means for constraints within an IMS to be unjustified and hence determine how this problem may be overcome. First, however, we will expand on our above question: is skill necessary for musical expression?

6.1 Is skill necessary for musical expression?

> This is a very exciting time for music. With the advent of computers, many of music's past restrictions can begin to fall away, so that it becomes possible for more people to make more satisfying music, more enjoyably and easily, regardless of physical coordination or theoretical study, of keyboard skills or fluency with notation. This doesn't imply a dilution of musical quality. On the contrary, it frees us to go further, and raises the base-level at which music making begins. It lets us focus more clearly on aesthetic content, on feeling and movement in sound, on the density or direction of experience, on sensuality, structure, and shape—so that we can concentrate better on what each of us loves in music that lies beyond the low level of how to make notes, at which music making far too often bogs down. ...For those who have wanted to do music but have lacked the background, computer intelligence may make it possible. (Spiegel 1988, p. 3)

There is a notion hidden in the above quotation, taken from the manual to Spiegel's Music Mouse software, that there is pleasure to be had in creating music without being particularly involved in how it turns out. Perhaps there is creative pleasure in hearing music that aligns with your 'intent'. Maybe just the sensation of being in control, or knowing that this experience is unique to you makes hearing music more rewarding. Or is the real enjoyment had simply through marvelling at the system, at the cleverness of its creator to have conceived of a musical world for us to explore?

Jordà (2002) pondered a similar question in the context of learning musical instruments, arguing that expressiveness does not necessary imply difficulty. We saw Jordà's (2004) model of the learning curve in Section 2.2.12 comparing the piano, violin, kalimba and kazoo. The kazoo may facilitate greater complexity early on but its expressive potential quickly levels out.

Jordà's model attempts to be neutral of musical genre or taste. But can we really assess the expressive potential of an instrument independently of the thoughts, ideas, prejudices and potential interpretations of those in the audience? Is he really justified in considering the output complexity of the piano to be vastly in excess of the kazoo? The piano may be polyphonic, but the kazoo carries micro, mid, and macro diversity when it comes to pitch, dynamic and timbral variation in (Jordà's own assessment of the key instrumental possibilities). The learning curve model relies on an assumption that how expressive one may be with a musical instrument is independent of how much time and effort has gone into mastering it. But is a kazoo virtuoso quite so inconceivable? Or is the kazoo a victim of Rokeby's operational clichés (Section 2.1.5), its lowly status on the musical instrument playing field *caused* by its low entry fee (both in time and money)?

Jordà's consideration of an instrument's possibilities independently of genre or performance context relies on our *classical* notion of expressivity (see Section 2.2.8)—and it is certainly a useful model within that context. This way of thinking is perhaps most apparent in Dobrian and Koppelman's consideration of virtuosity as being a necessary precursor of musical expression as 'the mind has more freedom to concentrate consciously on listening and expression' (Dobrian and Koppelman 2006, p. 279, see also Section 2.2.9). In other words, we enjoy watching virtuosi because they sound (and perhaps look) better. It is a reasonable enough notion but perhaps we are not entirely honest with ourselves if we imagine other factors do not enter into our assessment. Wiggins (2009) has argued that music is a psychological construct and therefore cannot be understood without considering perception. With an ecological consideration of expressivity, we must question the extent to which the expression of a performance is determined by the attitudes and expectations of those listening as well as the sound it produces. More bluntly, is anyone interested in listening to music that *anyone* could have made?

To answer this question, we return to consider the role of DMIs.

6.1.1 Why don't people like watching DMIs?

In Section 3.1 we identified that a large body of research into DMIs is motivated by a sense of frustration that they have not been accepted into the concert arena. Such research often assumes that DMIs, if created properly, might be able to fulfil the role of traditional instruments within a performance context (see Section 2.2.2). However, this does not seem to be happening to much extent, as much of the continuing research reviewed in Section 2.2 attests. The turntables are often cited as the most recently successful instrument (e.g. Jordà 2004) but these, of course, are not a DMI. What is it about *digital* that seems to put people off?

The use of computers is of course prevalent in musical production, for example the sequencer, sampler, software synthesisers. But they have failed to make a comparable impact on live instrumental performance (see Section 2.2.2). Is it really that hard to create a digital instrument that compares to an acoustic or even electronic instrument? Let us consider some of the criticisms that DMIs often face.

Expressive capabilities. They cannot create suitably diverse or expressive sounds (Sections 2.2.8 and 2.2.11).

Usability. The desired sounds are not accessed easily enough (Sections 2.2.11 and 2.2.12).

Virtuosity. Nobody takes the time to master them (Section 2.2.9).

Visual aspects. The gestures that arise when performing them are lacking in expression (Section 2.2.14).

Transparency. It is necessary to understand the relationship between gesture and sound in order to appreciate what is being expressed (Section 2.2.10).

Learnability is certainly an important issue to address with DMIs. However, given the considerable time and difficulty is necessary to learn non-digital instruments (even recent additions such as the turntables) it is not a convincing cause for their limited uptake.

There may well be a case that many DMIs lack visually, sonically, or how these two correlate. But if all that mattered was what we saw and heard then why do we bother with the interactivity? Why not automate the expressive aspects and rely entirely on ancillary gestures, i.e. 'lip sync' to a perfectly crafted recording? Such fakery would certainly not go down well with the audience. For example, as Schloss (2003) writes,

> Tape music may have been boring to watch, but it was honest . . . Some pieces performed nowadays claim to be interactive, but in reality they are simply not finished yet. So the performance involves the "baby-sitting" and "knob-twiddling" we might see on stage that is so unsatisfying to watch. At least with tape music, we can concentrate on the music; there is nothing else to worry about. (Schloss 2003, p. 240)

How can a performance be *dishonest*? Intuitively, we feel that we know the answer: a musician pretends to perform live but secretly everything that is heard is prerecorded. For example, the scandal of pop stars lip-synching to a dubbed vocal track (Auslander 1999). But what is so dishonest about knob-twiddling that, as Schloss suggests, might make it worse than having no performer at all?

Perceiving causality between the actions of the performer and what is heard was discussed alongside transparency in Section 2.2.10. We then argued in Section 3.3.5 that the need to establish transparency was a hurdle for DMIs as it shifted audience focus from the music to understanding the workings of the instrument. But causality is a more explicit requirement. We saw in Section 2.2.14 the argument that perceiving skill is crucial to enjoying a musical performance. It seems that, above the purely phenomenological experience of spectating music, we need to *believe* that what we hear is really being created in front of our eyes.

Fyans and Gurevich (2011) highlighted how unfamiliarity with an instrument (and its associated practice) makes assessing the skill of the performer difficult, with participants resorting to cues from the performers such as confident facial expressions and how embodied their relationship with the instrument appeared (see Section 2.2.14). But there remained a prejudice against the DMI, the Tilt-Synth. Rich sounds were heard but simple interaction observed. Any skill is thus interpreted as an intellectual rather than instrumental activity, 'mere button-pressing'.

6.1.2 Acoustic instrument counterexamples and the double standards of audiences

It would seem that, however impressive the sight or sound of a musical performance may be, we need to see the physical labour that has gone into it, evidence that it is the cause of what we hear. The reasons underlying this seem to arise perhaps from an underlying bias in our perceptual system towards the physical world and a need for evidence of the effort and skill of the performer.

Meeting this criteria may be necessary for a successful DMI performance. But is it necessary for *any* musical performance? Let us consider a few counterexamples.

The pipe organ

The church organ has been identified previously as grounds for hope for DMI researchers by Hunt and Wanderley (2002) as—whilst the sound does arise from a physical origin—its nature is completely unrelated to the controller.

There is no energy continuum. Organs have been electrified since the early 20th century (Sefl 2006). Prior to this, another person (the bellows operator) would provide the energy.

Compared to many DMIs, the church organ does not provide much expressive subtlety. Where control is given over the volume, it is on a 'per-channel' (i.e. per-manual[1]) basis and controlled via a pedalboard. Neither timbre nor pitch may be continuously altered. Perhaps most shockingly of all (for DMI designers at least) is the latency, which can be up to 200 ms on a large organ (Miller 1968) and vary with the pitch of each note—20 times the amount Wessel and Wright (2002) feel acceptable for a DMI. It is exceptionally difficult to play. The instrument is modal, with rows of stops available to select the timbre of the different manuals. However, the impact of this on audience transparency may well be quite immaterial as the organist is often hidden away in a box, both player and controller completely unseen by the audience for the duration of a recital.

Whilst the organ is perhaps not the most popular instrument, its place as an instrument has never been called into question in the manner many DMIs have been. But the church organ serves a specific *purpose*. It allows a single musician to fill an enormous space with rich music for a religious ceremony. They were created not for audiences, but for congregations. We have been left with a legacy of organ music and a tradition of organists to perform it. Perhaps, given its religious history, it is an unfair counterexample.

The trumpet

Let us consider the trumpet. Seeing a virtuoso trumpet player certainly is impressive. But is the trumpet really transparent to the audience? As an audience member, we can see the effort of blowing hard but beyond that all we see is the movement of three valves. The crucial aspect of control taking place between the lips and mouthpiece remains invisible.

The gestures that we see do not provide evidence of causality. But of

[1] Church organs often include a number of stepped keyboards referred to as *manuals*.

course, we *know* that there is causality. How else could there be sound eminating from the instrument?

Is this the embodied knowledge of the audience? Creating even a note on the trumpet is difficult without some instruction. And we might imagine such an attempt would make one more appreciative of the difficulty in playing the trumpet. But I would argue that it seems somewhat farfetched to suggest that without such an experience, our ability to discern a skilful trumpeter still relies on an embodied knowledge of the means by which a trumpet works.

The piano

The piano is perhaps one of the most popular recital instruments. Pianists fly all over the world to perform. However, in the standard concert hall setup, often half of the audience cannot see the pianist's hands or the keyboard. They may perhaps see some of the larger ancillary gestures, and the pianist's face. Whilst some people may try to get tickets on the left side of the house, those on the right are unlikely to suggest that they have been robbed of the live music experience.

It would seem that, important though it is for a DMI to demonstrate causality, we are more generous with non-digital instruments.

6.1.3 'Instrumentality'

Auslander (2002) argues that the use of the word *live* to mean not recorded did not appear, as one might expect, at the same time as recording technology but with the radio. It was only with the radio that the difference between live and recorded media became indistinguishable to the audience and the need to be explicit arose. An aspect of communication that people felt was important—simultaneity—was no longer evident in the medium and thus new practices emerged to make it evident—in this case explicitly stating when a broadcast was live.

In a similar manner, I argue that *instrumentality*, the degree to which a DMI may be played in the manner of an instrument (see e.g. Cadoz 2009; Gurevich and Fyans 2011), was not something that needed to be considered with acoustic instruments. There was never any question as to whether a technology was an instrument as that was all that it could be.

Instruments exist primarily to create *sound*. Perhaps some designers considered how a performer might look when they played but this was surely a minor consideration. As we have seen, instrumental performance is about much more than sound. Otherwise, we would simply stay at home and listen to the recording. But, as Small (1998) points out, the fact that we can now listen to recordings has affected what we consider important in performance.

Causality is important, but before the technology arose to reproduce sound, it did not need to be explicit. With acoustic instruments, it does not need to be made apparent as it is self-evident. The button accordion player is not criticised for 'mere button-pressing'. Likewise, if an accomplished pianist hears after a performance that they make playing the piano look 'effortless', this is likely to be received as a compliment rather than a suggestion that they have chosen too easy an instrument. But now we have instruments that *can* make performance 'easy', removing the difficulties that were necessary to putting expression into a performance. It has become painfully apparent that musical performance is about more than hearing expressive music created in real time. We need to believe that creating it is difficult.

6.1.4 DMIs and the potential for automation

We saw in Section 2.2.9 Dobrian and Koppelman's suggestion that the elephant in the room of NIME performance is a lack of virtuosity. But an even bigger elephant is the fact that any sound coming out of a laptop could have been produced identically by a recording. In fact, sometimes recordings are involved. We simply do not know because there is not an established norm of how much automation is appropriate within a DMI performance. As a result, there is a loss of the concert 'pact' that allows unworried trust that what is being seen is genuinely the work of a skilled performer (see Section 3.3.2).

DMIs range from generic, open tools to highly idiosyncratic 'signatures' of a composer charged with their musical ideas (Magnusson and Hurtado 2008). By contrast, traditional instruments have clearly established (and mutually understood) boundaries (Schloss 2003). Magnusson (2010) argued that it is at the limits of an instrument that creativity happens. But if we in the audience do not know what those limits are then we are not sharing in the significance of what is happening.

We do not know the limits of what we are hearing is because *there are no limits* to what we can hear from an arbitrary DMI. The limits we are interested in are how much of a performance can be created live, and that is why evidence of causality is crucial. Without this, a sense of wonder is tainted with a suspicion of how real everything is—something we were able to exploit with the Serendiptichord (Section 3.3.2).

Schloss and Jaffe (1993) provide an example of the theatrical juggling troupe the Flying Karamazov Brothers embedding wireless MIDI transmitters in their helmets which they would trigger through juggling. But the audience simply assumed they were syncing to a recording. Schloss and Jaffe see this as reaching the threshold of 'magic'. Reeves et al. (2005) described *magical interaction* as that where the effects are revealed but the manipulations that caused them hidden (Section 2.1.7); Schloss (2003, p. 242) takes a similar meaning when he says 'Magic in a performance is good. *Too much magic* is fatal! (Boring)'. But magic denies rational explanation and is far from boring. When the sound has passed through a laptop there is always a rational explanation. Once *digital* is involved, effects due to invisible manipulations are generally assumed to be *automation*, not magic. It only becomes magical when *no* reasonable manipulation can be imagined. Without the concert pact (see Section 3.3.2), trust in interactivity must be either requested through explanatory talks or programme notes (e.g. Paradiso 1999) or earned through a demonstration that would be implausible if scripted (e.g. Saltz 2001). Hence the need for visible instrumentality in DMIs.

6.2 Skill and authority in live music

We have established thus far that it becomes harder to convince an audience of a skilful performance as technology capable of automation is introduced. Why do we need to convince them in the first place?

Understanding of authorship, intention and stylistic relevance plays a fundamental role in aesthetic appreciation (Leder et al. 2004). As Cascone (2003) argues, the demonstration of musical skill provides the performer with the authority to be listened to.

The need for musical authority is sometimes cast under a cynical light (e.g. Auslander 1999), especially if we adopt the viewpoint of classical expressivity that we are here exclusively for a sensory experience. However, under more ecological considerations of musical performance (see Section 2.2.8), an understanding of origin may well play a fundamental part of the musical experience. As one respondent of Lindström et al.'s (2003) questionnaires on music students' attitudes wrote: 'expressivity comes from within your soul'. It is surely not that unreasonable for our listener to wish to know whose soul that might be.

Placing a significance on the origin of music is in line with the communicative aspects of musical perception (Section 2.4.2)—but also the expectation-oriented models (Section 2.4.3). Listening to music is an active process. Musical authority can establish the expectation that what we will hear is worth the effort of listening to with an open mind.

Many DMI performances are a single individual's work as instrument creator, composer and performer (Magnusson 2010). The problem is not that anybody doubts the performer's authorship of what is heard. But creating music live presents a particular challenge: there are no second chances and there is no time to stop and think. To give Clarke's (2006) example, listening to the soloist of Sibelius's violin concerto would be a very different experience if we thought they were improvising.

The problem with 'mere button pressing' is specific to DMIs: we cannot tell if we are hearing something that was prepared earlier. Therefore the performer cannot demonstrate the skill necessary to establish authority. This *is* a problem of audience familiarity as the degree to which the performer is presently involved is more apparent with an understanding of the mapping. However, it is not a problem that acoustic instruments (or indeed, electronic instruments) have to deal with.

Live music holds a special place for many music fans but we often seem to have difficulty expressing quite why. We are not suggesting that watching a visible demonstration of skill is the sole reason. However, in the familiar performance context where an audience has arrived to listen to a performer, the musical authority of the performer plays a role in appreciating the music that they create. A demonstration of musical ability is one means by which this authority may be established. (Other means may include, for example, a positive review from critics and friends.)

6.2.1 Technique, expression and creativity

There are undoubtedly a large number of skills demonstrated in musical performance. However, there are three in particular that will be relevant for the framework we develop below: *technical*, *expressive* and *creative*. These will allow us to demonstrate the different expectations provided by different performance contexts.

We saw in Section 2.1.8, Sheridan's (2006) distinction between *technical* and interpretive (i.e. *expressive*) behaviour in performance. This distinction has also been drawn in music by Godlovitch (1998). Auslander (2005) further argues that a fundamental difference between human and machine performers is that humans have both technical and interpretive skills whereas machines have only the former. Similarly, Lindström et al.'s (2003) aforementioned study found expressivity is commonly considered to distinguish great musicians. Technique is good, but not enough. This distinction is also mentioned by Schloss (2003, p. 242): 'people who perform should be performers'.

However, both technical and expressive performance require material to perform. So that we might draw a distinction between improvisation and recital we will also consider *creative* skill—the ability to come up with original material. Creativity has been demonstrated by Eisenberg and Thompson (2003) as forming an important part of how individuals evaluate improvised music.

How does this relate to interactive music? In our previous study, we found some participants apparently frustrated by their inability to draw on their own skills. At the same time, others interpreted the system in a different manner and were content to explore it as a demonstration of another person's musical ideas. If, as I am suggesting, the demonstration of skill is necessary in a musical act then it is important to understand how an IMS may facilitate this. To do so, we need to understand what kind of skills are necessary, who may provide them (e.g. the participant, the system) and how this may be demonstrated to the participant.

In the following section, I will introduce the concept of agency, which will provide a more specific means to understand what it means to demonstrate skill. Following this, I will introduce a framework of Perceived Agency that will allows us to formally categorise the different types of skill, how they are demonstrated and to whom they are attributed.

6.3 Agency: The freedom to demonstrate skill

I argued above that the demonstration of skill is an important part of musical performance in need to consideration. In this section, we will meet the concept of *perceived performative agency* as the perceived bounds within which that skill may be demonstrated. We will then be able to distinguish different performance contexts based on the different types of performance agencies they infer and identify how problems may arise if this is not perceived to be the case. From there, we may then consider the consequences for the interactive domain.

First, however, we will formally define what we mean by ‘agency’.

6.3.1 Formal definition

Agency describes the ability of a person to take actions, have initiative and influence outcome (Tanaka 2006). It arises in a number of fields including artificial intelligence, cognitive psychology, moral philosophy and cultural theory with slightly different formulations. An individual acting with agency is described as an *agent*.

We will build on Bandura’s (2001) definition of personal agency from social cognitive theory. A human is acting with personal agency when the following is satisfied.

Intentionality. Intentions are plans of actions with predictions about their outcomes. Plans need not be specified in full and the agent is proactively committed to bringing them about.

Forethought. The agent is able to create representations of future events that inform planning. Anticipated events may be then be established as goals. Forethought includes representing conditional relationships between environmental events and how actions will affect these outcomes.

Self-reactiveness. The agent motivates and regulates themselves. Having established a plan, they cannot simply sit back and await the outcome but must be able to be involved throughout in order to regulate its execution.

Self-reflectiveness. The agent is able to reflect not only on events and actions but on their own functioning. A belief in one’s own efficacy is fundamental to agency. Without a belief that one has the power to produce desired effects, there is little incentive to act regardless of goals or motivation.

(Bandura 2001)

Note that choosing *overall* goals is not necessary. We may have no choice about whether we need to eat, but may still establish goals in order to satisfy this overall need. From this we may arrive at a notion of degrees of agency or, more specifically, defined parameters within which an agent acts.

6.3.2 The importance of agency

Agency is a key issue to consider with Interactive Music Systems. How much freedom (and hence, responsibility) do we grant our user? Are we providing a context for them to discover their 'creative side'? Or are we simply constructing an experience for them to participate in?

Whilst recent years have seen much debate on the role of copyright in today's world, the importance given to correctly attributing work remains as high as ever. Originality—by which we mean uniqueness—is important to musicians (Magnusson and Hurtado 2008). But it is crucial that originality is identified to arise from a musician. Thus, a digital tool that seems to impose another's ideas of how music should be made is seen as an unwelcome limitation (a loss of originality) (Magnusson and Hurtado 2008). However, the commonly understood limitations of a musical instrument provide the structure in which a musician may develop their own style (Wallis et al. 2011; Gurevich et al. 2009).

6.3.3 Perceived performative agency

Our motivation for considering agency arises from an audience being able to attribute whose demonstration of skill has contributed to a performance. However, we will tend to avoid the terms *authorship* or *attribution* as we wish to emphasise that perceived agency is something that is perceived *intuitively* and *subjectively* rather than socially negotiated or explicitly declared. An individual may declare themselves 'in control' and there may be an accepted author. We are considering how far this is felt to be true by an audience.

Agency has previously been considered within the context of Digital Music Systems (DMSs) by Tanaka (2006) in terms of a participant's sense of their own actions within the music produced through their interactions with a DMS or within an ensemble. However, the framework presented here extends this notion through considering agency over different aspects of a performance, the differing ways and roles in which it may be realised and how it is conferred by a performance context.

Informally, we may think of performative agency as being attributed to those who are responsible for making a performance event what it is. We define the *perceived performative agency* of an agent over an aspect of a performative event as *the degree to which the aspect of the event is perceived to have resulted from the agent acting with agency*, where *agency* is as defined in Section 6.3.1. For brevity, we will use the term *perceived agency* to mean *perceived performative agency*.

Similar to Tanaka's formulation, perceived agency is entirely a *subjective* phenomenon of the perceiver. There is not a single objective agency of which a spectator may have an 'inaccurate' subjective interpretation. For simplicity, we will often refer to agency rather than perceived agency. However, by this we are always referring to the representation that an individual has imposed upon the performative event.

An individual infers from a performance context *norms* of who should be exercising agency, how and in which role. For example, a classical recital does not usually grant the performer agency over which notes are to be performed. However it does grant them agency to perform the given notes through their own technical expertise. Note that it is not possible for an agent to 'refuse' agency inferred by the performance context. Granting agency is akin to specifying a set of available actions from which the agent may then choose. It is both the provision of freedom and the assignment of responsibility.

Therefore, in order to perceive the agency of another individual it is necessary to perceive that their actions have been chosen rather than determined by other factors, which in turn requires the perception that there were other actions available. In this way, the perception of agency requires an understanding of what *could* have been done but was not. We saw a similar notion in Section 6.1.4: without an understanding of the limits of a DMI, we cannot perceive when they are being approached by the performer.

The condition of self-reactiveness (see Section 6.3.1) means that once an outcome has been fixed, then the performer is no longer exercising agency. A process that was prepared and begun earlier *involved* agency but it only *involves* agency to the extent that the agent is involved in its execution. Thus agency is exercised over a period of time. This notion will allow us to distinguish below between *live* and *preparatory* agency.

6.4 A framework of Perceived Agency

We are now ready to present the framework of Perceived Agency. The framework is not intended to be a theory of origin that intends to explain why we enjoy live music. Rather, it is presented as a tool to identify how individual roles are inferred and acted upon by different performance contexts, as we shall demonstrate below. It is through examining how musical performances are perceived that we may understand how problems might arise within new performance and participatory contexts.

Although we will consider three types of performative agency, corresponding to the three musical skills discussed in Section 6.2.1, this is not to suggest that the demonstration of other skills is not important within music. For example, we might also consider passion, leadership or ensemble communication. However, as we shall see in the below applications of the framework, these three arise naturally when comparing different contexts of performance.

6.4.1 Types of agency

We consider three types of distinct contribution in which agency may be perceived at a performative event: *technical*, *expressive* and *creative*, which we denote as T, E and C respectively.

Note that we are not claiming that these are necessarily distinct ingredients that form a musical work (indeed creativity may well arise from technical ability). However, based on the literature reviewed in Section 6.2.1, we argue that these aspects may commonly be *perceived* as distinct qualities in a performance. The extent to which they are perceived independently may vary across individuals.

Technical agency (T)

Technical agency (T) is exercised with the application of skill, both technical and abstract, to realise an outcome. The skill may be embodied or abstract. T might also be described as executive agency.

T may be considered loosely associated with virtuosity (Section 2.2.9). However, we may also imagine a performer to be described as 'technically accomplished but lacking in emotion or imagination' to mean that T was perceivably demonstrated but not our other types of agency outlined below.

(It is beyond the scope of the framework, but we might also imagine a sporting performance as consisting mostly of T.)

Expressive agency (E)

We define *expressive agency* (E) as the ability to feel, communicate and interpret emotion. This includes interpreting the sentiment of a score, realising it through expressive performance. We may consider demonstrable emotional intelligence as a part of E. It also includes *having* emotion to communicate.

In Section 6.2.1 (as well as Section 2.2.8) we touched upon arguments about whether machines can feel or express emotions. This is not a debate we need worry ourselves with. Agency is *imposed* upon a performance by the perceiver (see Section 6.3.3). An important consequence of this, however, is that *thinking* a performance is being rendered by a machine may reduce the potential to perceive expression.

Performances without a perceivable demonstration of expressive agency might be described as 'mechanistic', 'lifeless', or 'soulless'. Recall, however, that agency may be exercised within an externally defined goal. Musicians in a symphony orchestra may not determine the expressive content of what they perform but still exercise E in interpreting the conductor's expression, understanding it and then realising it in their performance (cf. Small 1998).

Creative agency (C)

Creative agency (C) is the ability to think up original (and good) content. We may also describe it as having imagination or vision. The actual *realisation* of these ideas rests with E and T. A demonstration of C without E or T might be described as 'good ideas but poor execution'.

6.4.2 Live and preparatory agency

We define a type of agency as being perceived *live* if this period of time is concurrent with the performance and *preparatory* (abbreviated as prep) if it took place before the performance. Agency may be perceived as both preparatory and live. For example, an awareness of the amount of rehearsing that has gone into a recital may be perceived as both live and preparatory technical agency.

It is my position that an assumption that live creation is a more skilful equivalent of prepared creation is somewhat simplistic.[2] However, this is not dealt with by the framework because agency is imposed subjectively on a performance by the perceiver. Therefore, discussions of what kinds of skills are more important or difficult remain subjective to the individual.

[2]For example, Johnson-Laird (1988) argues that composition and improvisors are distinct skills evidenced by the existence of improvisors who cannot compose and composers who cannot improvise.

However, the framework does highlight how different performance contexts imply different norms about whether live or preparatory agency is to take place. We might therefore speculate that individuals choosing to attend a particular performance context will place a greater value on the types of agency its norms imply.

We shall notate the *preparedness* of agency type A as A(live), A(prep) or A(live, prep) for live, preparatory or both, respectively, where A may be one of T, C or E as defined in Section 6.4.1.

6.4.3 Agent roles

A musical performance will often be perceived as arising through the actions of a number of agents. We categorise these into three *roles* of the listener, performer and other contributors corresponding respectively. These are analogous to the grammatical notions of first, second and third person that might be used if the perceiver were conversing with the performer and will therefore be denoted 1, 2 and 3 respectively. The framework describes how one individual perceives agency of an event. This perceiver is therefore considered in the first person ('I'). The focus of the perceiver's attention is considered in the second person ('you') as they might in a conversation. Others are then considered as third parties ('he', 'she', 'it', 'they') We will discuss second and third person agency first before moving onto the special case of first person agency.

You: performers (2)

Within this framework, performers are those that are visible and in receipt of attention by the audience. They therefore correspond to the second person ('you') and are denoted 2.

Note that although we might consider unseen parties (e.g. the sound and lighting technicians) as 'performing', unless they are the focus of attention (e.g. on stage) then they would not be considered as performers within this framework. Performers are often easily identified as those on a stage.

They: non-visible contributors (3)

Non-visible contributors may include the composer, director, choreographer, producer, curator and those backstage such as a sound technician. They correspond to the third person ('he', 'she', 'it' or 'they') and are this role is therefore denoted 3. Even if, say, the composer were present in the audience, we would still consider them a non-visible contributor.

I: the audience member or participant (1)

First person agency is primarily how far the audience member is instrumental in bringing about their experience. The audience member is the individual around whose subjective perception the framework is considered. Therefore, they correspond to the first person ('I' or 'we') and are denoted 1.

We may imagine our T, C and E types of agency playing a part in this, for example as follows.

Technical agency: *I am clever for being able to appreciate this.*

Creative agency: *The beauty of this touches me in a personal way (i.e. others may not see the beauty in this).*

Expressive agency: *I connect with this work on a personal emotional level.*

Whilst we argued above that listening to music is an active cognitive process (Section 6.2), we may also consider perceived first person agency as the degree to which the listener feels *ownership* over the experience. By this we mean how *unique* the connection between listener and performance is (in contrast to other people). It is arguable in such a case whether the formal requirements of agency defined in Section 6.3.1 are met in this case. A performance may have been arrived at by chance rather than by intention but still feel strongly personal and unique. Therefore, we will refer to this case of personal uniqueness as a 'looser' definition of first person agency.

The company of others affects the experience of listening to music and we might imagine our framework extending to consider the shared experience of the audience ('we'). In this case, the uniqueness would apply within the group ('we are making this experience what it is'). However, a consideration of the means by which musical events binds groups of individuals together is beyond the scope of this thesis.

6.5 Perceiving agency in established performance contexts

In order to demonstrate the important role that attributing agency plays in the aesthetic experience, we will in this section apply the framework to established performance practices. As we shall see, whilst there is considerable variety in the types of agency exercised, the *importance* of perceiving agency remains constant. Technology that impedes this process has the potential to negatively affect a performance. Therefore, agency is as much an engineering issue as it is artistic or cultural.

Our motivation is to identify potential problems with new musical contexts and not to make value judgements on existing practice. The framework is based on observation of existing performance practices. Therefore, to apply it as a means to then assess the value of these practices is somewhat misguided.

It is important to note that the framework does not describe the musical *quality* of any aspect—i.e. how good it sounds—nor how much skill might be required. The perception of agency may inform the extent to which a particular performance is perceived as skilful, but the manner in which it does so remains subjective to the individual spectator.

In particular, there is a danger when considering agency to conclude that the amount of agency attributed to a performer is somehow indicative of their talent or that performers without a particular type of agency are somehow 'subservient' (e.g. Small 1998). It is possible that the raw musical signal produced through live improvisation would not be judged as 'accomplished' as one that was composed and performed. Or vice versa. But with a greater appreciation of the role of perceiving agency we may see that drawing such a comparison without considering both performance context and the listener is not particularly valid or useful.

6.5.1 Classical concert

A classical musician will typically demonstrate live and preparatory technical and expressive agency: 2T(live, prep) and 2E(live, prep). We may expect a recital musician or conductor to have more E than an ensemble musician, although orchestral musicians still demonstrate E as they are required to interpret expressive instructions into expressive performance. Classical musicians usually perform music composed by another and therefore are not typically attributed creative agency (C) whilst performing. However, on occasion we may imagine the perception of C(prep) for an especially creative interpretation of a work (e.g. Glenn Gould's performances of Bach's Well-Tempered Clavier).

The composer (3) has C, T and E all as preparatory agency. Here, C is the imagination to come up with musical ideas, T is the skill in scoring those ideas and E is the ability to embody expressive content in the work. We may also imagine the perception of 3E being enhanced through a belief in the passion of the composer.

The listener (1) may also feel that their experience and understanding of the genre is contributing to their experience, providing T(live). They may also feel they are being creative in their appreciation (C) and personal connection to the expressive content of the music (E). We may imagine there being a greater amount of live 1E if the listener feels themselves forming of affirming a personal connection with the musician, composer or sentiment of the music whilst listening.

6.5.2 Jazz improvisation

With improvised jazz, the performer (2) is expected to exercise C(live). This then allows for a great deal of T(live) and E(live) to be attributed. In contrast to classical music, the composer (when we are hearing a 'standard') is less prominent, thus for 3 we have less C, T or E.

The use of standards within jazz allows for a clear distinction to be made between 2C and 3C. We may also observe musical ideas being passed between players within an ensemble providing evidence of C(live) rather than C(prep).

As with classical music, the jazz listener may feel a personal level of intellectual and creative input goes into appreciating what they hear. However, with the perception of 2E being live to a greater extent than classical (and significantly less 3E), we may find a greater level of 1E(live) through a personal connection to the performer.

6.5.3 Indie gig

We will consider an 'Indie' gig as a relatively small performance by a rock band that are not generally considered by those attending as 'mainstream'.

Lingel and Naaman (2012) presents evidence that fans of indie music often feel a personal connection with the band and that being a fan may be felt to be an act of personal expression, providing 1E. We may further draw from Auslander's (1999) discussion of the importance of 'authenticity' in rock music that it is important that the emotional content of the music is felt to arise genuinely from how the musician feels, making 2E important. In order to maintain this perception of 2E, it is therefore important that the expressive content of the music is not perceived to have arrived from a third party (no 3E). As a result, 3C and 3T are generally disliked as well, and indeed we may see the strong belief that a performer should write their own music as a means of providing evidence of 2 as the origin of E through 2C and 2T.

Therefore, indie performance norms typically demand a large amount of 2E(live) and some amount of 2T(live) but at the same time 2E(prep), 2T(prep) and 2C(prep). The extent of third party involvement is minimised. 3T is permitted provided it is thought not to provide 3E or to detract from 2T. Thus, mastering a record is acceptable whereas using AutoTune to disguise poor singing is not.

6.5.4 Differing norms of perceived agency

We summarise the above observations in Table 6.1. Note, however, that these observations are not presented as what actually happens or necessarily what is perceived. They are our observations of what is considered important within different musical paradigms and therefore indicate how performance contexts may invoke these paradigms to infer *norms* that agency will (or perhaps *should*) be distributed in this manner.

Table 6.1: Observed norms of perceiving agency in established performance contexts.

Performance paradigm	Expected perceived agency		
Classical concert	1T	2TE(live)	3TCE(prep)
Jazz improvisation	1TE	2TCE(live)	
Indie	1E	2TE(live)	2TEC(prep)

Where types of agency are merged this implies both types in the given role, e.g. 2TE(live) should be interpreted as 2T(live) and 2E(live)

If a performance context implies a performance within one paradigm and the norms of agency of this paradigm are not perceived, then the music may be considered 'inauthentic'. For example, a bebop performance in a busy basement jazz club may imply the paradigm of jazz improvisation. If an audience member observes that the performance was note-for-note identical to a previous performance by this musician, then they may perceive 2C(prep) instead of the expected 2C(live) and object that this was not 'authentic' bebop. If they recognise the performance as being note-for-note identical to that of a different performer then they may perceive 3C(live or prep) instead of 2C(live), and again feel the performance to be inauthentic. However, the same audience member may later find themselves at a recital by a classical musician performing a piece note-for-note identical to many performances earlier and take no exception to this fact.

In this way, we may see that whilst there is not a universal distribution of agency expected from a musical performance, a performance context may create expectations in the audience of a specific distribution and a perception of *inauthenticity* if this is not met.

6.6 Perceiving agency in new performance contexts

We will now apply our framework to consider how an audience perceives (or struggles to perceive) agency within new performance contexts. We will not necessarily consider every role for each context but highlight distinctive aspects. For example, most of the new performance contexts we consider are quite niche and therefore their audiences may feel a sense of personal agency.

6.6.1 DMI performance

Returning to our opening topic of Digital Musical Instruments (DMIs), in the context of agency we can see that a difficulty often arises in distinguishing between (live) and (prep). DMIs are often performed in a classical concert context which would imply the need for 2T(live) and 3E(live). But as there is not an established norm of how much material is prepared within the DMI, this may not be assumed through the concert pact (see Section 6.1.4). Therefore, evidence is required to make it clear that these agencies are being exercised.

However, there is scope for additional agency within a DMI performance. The instrument itself may be seen as the result of technical and creative skill prior to the performance, allowing the instrument designer to be attributed 3T(prep) and 3C(prep). There is also additional potential for 1T(live) as the audience may feel they are demonstrating technical ability as they develop an understanding of the mapping. However, in contrast with composed instruments, with DMI performances, this role is typically sought to be minimised as it is typically necessary to understand the mapping to perceive the live agency of the performer (see Section 6.1.1).

The difficulty in distinguishing between preparatory and live agency may also help explain why DMIs are often only performed by their creators (see e.g. Fels 2004). When the creator is performing, the difficulty is in distinguishing between 2(live) and 2(prep). However, if the instrument is created by another person, then this becomes the difference between 2(live) and 3(prep), potentially limiting the performer's ability to demonstrate agency.

6.6.2 Generative music: machine agency

Generative music—that created by a computer without any input—can be either performed live or rendered offline with a recording presented. We shall consider performative generative music here.

Can a machine have agency? From a philosophical viewpoint, Auslander (2005) argues that they can if they are acting as performers and not following a script. The framework is applied subjectively, so we do not need to worry about whether a machine can 'really' have agency. However, it is reasonable to infer from Auslander's comment that a viewer may *perceive* a machine as having agency. This point is relevant with regards to self-reflectiveness (Section 6.3.1), the requirement that an agent to believe in its own power. Whilst belief is an ambiguous notion with regards to a machine, we may perceive a machine to be acting as if it believed in the power of its actions.

Therefore, it is reasonable to consider that a generative system may be perceived as exercising live technical and creative agency, 2T(live) and

2C(live). However, as with other computer-oriented systems, without evidence of its generative capacity, it may be difficult to distinguish between 2(live) and 3(prep). Without input, providing conclusive evidence of live agency is difficult. However, we may still perceive agency if this seems more likely than a recording. This may be based on how 'unhuman-like' the output of the system sounds what we are told about the system.

A key difference between machine and human agency is that machines are typically perceived as having been created by a human. Indeed, the output of a generative system is generally attributed to its creator (Ariza 2009). Thus, the agency exercised by the machine may well be seen as a product of its creator. In this way, demonstration of live agency by the system is simultaneously a demonstration of preparatory technical and creative agency by its creator. Agency is perceived at two complementary levels.

6.6.3 Rowe's Interactive Music Systems

Recall form Section 2.2.4 that Rowe (1993) defined three aspects by which to classify his Interactive Music Systems:

- score or performance driven in input,
- sequenced, transformative or generative in output[3] and
- the instrument vs player paradigm.

Rowe's description of a sort of artificial player implies an expectation that the system will be interpreted as its own agent. Thus we have two performing (2) agents. Let us refer to them as *musician* and *system*, which we shall denote as (2_M) and (2_S) respectively.

Under our framework, this third aspect may be seen as the difference between the output of the system being perceived as a demonstration of the musician's (2_M) agency (under the instrument paradigm) or the system's

[3]Note that the *input* and *output* clarification was our interpretation and not explicitly defined by Rowe (1993).

(2_S) agency (under the player paradigm). Systems falling under the instrument paradigm we may consider as analogous to a DMI and hence defer to Section 6.6.1. We will therefore consider only systems within the player paradigm here.

The first of the above aspects distinguishing between score or performance driven may be perceived as where the balance lies between whether the decision of when to play was determined in advance by the system's designer or is to be determined live by the musician. This leaves this aspect as a spectrum of where creative agency is perceived with 3C(prep) describing the score-driven system and 2_MC(live) the performance-driven system.

In a similar manner, the extent to which the output of the system relies on prewritten material may be seen on a spectrum between 3C(prep) and 2_SC(live).

In both of the above distinctions, the musician performing with the system may also be known to be its designer, which would make 3 and 2_M the same person.

Note here that the live agency of one performer allows the live agency of another to be perceived. We saw a similar example above with the jazz group where bouncing ideas evidenced the improvisatory (i.e. live) agency of the performers.

We may consider the Continuator (Section 2.2.4) as an example. The Continuator (I would argue) is likely to be perceived as an agent embodying 2_ST(live). This agency is then attributed to its creator as 3T(prep) (*that's clever*) and 3C(prep) (*an imaginative idea*). Of course, it would not be as impressive if we believed its responses to be preprogrammed (3T(prep) instead of 2_ST(live)). But, through the interactions with the (improvising) musician, the system's agency is demonstrated to be live.

6.6.4 Composed instruments and the Serendiptichord

The composed instrument, in the form presented in Section 3.3.1, is similar to the DMI in that the distinction between preparatory and live agency is ambiguous unless causality is explicitly established. Forming an understanding of the mapping forms a part of the audience experience requiring 1T(live). I would argue that a strong motivating reason for the audience to form this understanding is the desire to distinguish between 2(live) and 3(prep).

We exploited this desire with the Serendiptichord (Section 3.2), maintaining ambiguity of the extent to which agency was 2(live) or 3(live or prep), i.e. whether the instrument was interactive or whether the sound was either a recording or controlled backstage. Furthermore, over the performance the instrument began to project an impression of agency itself which, using the notation of Section 6.6.3, created ambiguity between 2_M and 2_S. By presenting the instrument in a more theatrical manner rather than that of a classical concert (Section 3.2.5), we explicitly avoided the concert pact (Section 6.1.4). This allowed the potential use of preparatory agency to be presented as an implicit challenge for the audience to work out rather than a potentially distracting concern that the pact was being violated.

6.6.5 Live coding

In live coding performance, the performer's code is usually made visible to the audience often with comments to aid audience understanding (Section 2.2.4). This gives a strong 1T(live) agency through the audience's attempts to decipher the code (often drawing upon specialist knowledge). The projection also serves to demonstrate the live agency of the performer. It not only allows causality to be inferred but provides an unusual insight into the performer's thinking process demonstrating some of the more specific attributes of agency (e.g. self-reactivity). The use of previously defined procedures within live coding also allows for preparatory agency to be introduced explicitly (and therefore 'honestly') without disrupting the perception of live agency.

The above analysis demonstrates that whilst the distribution of agency is less established within new performance contexts, perceiving agency remains important whether or not it is explicitly considered by performers or spectators. In the case of the Serendiptichord, we have seen how uncertainty over the norms of agency may then be exploited in the performance of a composed instrument.

Table 6.2 shows an illustration of how a spectator may perceive agency from the different contributors involved in a musical performance. Observe in this table that whilst the sound engineer may need to be creative and technically skilled in how they balance the different audio channels on a mix, in this illustration they are not *perceived* to be exercising agency. Perceived agency is not an objective reality of who does what but of who is perceived to be responsible for it.

6.7 Perceiving agency in interactive music

As we consider participatory interactive music paradigms, the boundary between 1 and 2 may seem blurred as the audience may also be considered the performer. However, as we have seen so far, listener agency is common throughout all performance contexts. Thus, the framework may be applied most consistently by keeping the participant, which we shall denote P as 1 (the perceiver of agency), the system, if perceived to have agency as 2 and its designer as 3.

As we are only considering interactive music that does not require non-participant input during an interaction, 3 is always preparatory. Where other participants are involved, then they will be 1 if collaborating with P, 3 if spectating and 2 if they are participating but P is not.

Table 6.2: How the agency of different contributors to a musical performance might be perceived. Note that this is an example of how one individual might perceive agency not a statement of how far these roles actually contribute or demonstrate skill.

Role	**Perceived agency**
Recital instrumentalist	2TE(live, prep)
Orchestral instrumentalist	2T(live, prep), 2E(live)
Improvising instrumentalist	2TCE(live)
Indie performer	2E(live), 2TCE(prep)
Offstage singer	3TE(live)
Conductor	2E(live, prep)
Studio composer/producer	3TCE(prep)
DMI designer	3T(prep)
Composed instrument designer	3TCE(prep)
Generative music creator	2TC(live) implying 3TC(prep)
Rowe's artifical performer	2_STC(live) implying 3TC(prep)
Sound engineer	(none)

Where types of agency are merged this implies both types in the given role, e.g. 2TE(live) should be interpreted as 2T(live) and 2E(live)

6.7.1 Reactive music

Reactive music systems respond to environmental input that was not typically intended as input (see Section 2.2.5). This provides evidence that any agency being perceived is live rather than preparatory. In this sense, we may see reactive music as akin to generative music with the addition of evidence.

There is a further consideration to reactive music. Reactive music responds to environmental consequences of the *listener*'s actions (e.g. sounds from the place they have chosen to be, their subconscious bodily movements). We might consider cases where the listener is acting with agency over what they hear such as choosing environments with interesting sounds as having listener agency of our familiar types (perhaps 1C and 1T). But even without this intention, the listener may still feel a sense of personal ownership of what they hear, which fits within our looser definition of perceived (performative) agency (Section 6.3.3).

6.7.2 Interactive music systems

With interactive sound installations, the participant has typically not rehearsed or done any preparatory work prior to the interaction, meaning that there is typically no 1(prep). In contrast, with interactive sound applications this may not be the case (see Section 2.2.3).

A potential participant may not be expecting their agency in the experience to extend beyond their interpretation. The first challenge of the system is often to communicate to that their agency extends beyond this through establishing a perceptual feedback loop. Without this, the system is subjectively equivalent to a generative or reactive system.

As the participant develops a mental model of how they may interact with the system, they are exercising 1T(live). This interaction may provide the participant with evidence of the live agency of the system, 2(live), which, as with generative systems, may likely lead to 3(prep) agency. Some designers may also wish for their participants to feel a sense of 1C(live), 1E(live).

Well, this may be what the system's designer had in mind. However, the participant may perceive the context as requiring more 1T or 1C than they are prepared for (*I don't know how to use it* or *I don't know what to do* respectively). They may feel that the system provides 1E and feel uncomfortable exposing themselves in front of others. On the other hand, they may reach a stage where they feel bored with the limits of 1T (*I've learned all there is to know*), disappointed with the limits of 1C (*I don't have enough control over the output to create anything that is mine*) or 1E (*the sound doesn't speak for me*).

As we can see, similarly to DMIs, the norms of interactive sound installations are not exact in determining where agency should be perceived.

Thus, difficulties may arise when participants feel entitled to more agency (or intimidated by the degree to which they have been granted it).

6.8 Discussion

Perceived performative agency defines the structure in which skill may be demonstrated. The perception of agency is different from authorship. It is not just 'who did it?' but 'what was the structure in which they did so?' Perceived agency is what allows us to discern how much skill was demonstrated given an outcome. A key conclusion is that the perception of skill, creativity and expression depend as much on what *could* have been done as it does on what was done. The idiomaticities of an instrument define not only its voice (Tanaka 2006) but a space of potential musical acts. Audience understanding of this potential allows the significance of the one performed to be understood.

We may now return to the question we asked at the beginning of this chapter: does skill matter? The performance contexts considered suggests that a demonstration of skill does matter. Although there is diversity in terms of what skills are demonstrated how and by whom, it remains important that it is done so within the scope for agency implied by the norms of the performance context.

Without established norms of agency within interactive music, we have not ruled out the possibility of creating an Interactive Music System (IMS) that does not require skill of its participants. However, problems may arise when participants infer expectations of agency that are then not met by the system. For example, assumptions by a participant that a system exists as a means for them to be creative may make them unappreciative of any musical ideas embedded within the system—regardless of the quality of those ideas. Conversely, expectations of discovering music within a system may make a participant unappreciative of an interaction in which they must exercise agency in order to determine an outcome. Both of these cases represent a mismatch between the designer and the participant over the balance between 1C(live) and 3C(prep). In either case, *someone* needs to take responsibility for what is heard. But this suggests that it is not enough simply to demonstrate agency; it needs to match the participant's expectations.

Whether or not the participant demonstrates skill, the framework suggests two reasons why interactivity may still enhance the appreciation of a work. First, interactivity provides the means to establish whether agency is live or preparatory, for example with a reactive music system (Section 6.7.1). Note though that such a demonstration is possible without the participation of the perceiver as we saw, for example, with Rowe's artificial performer (Section 6.6.3) or musicians within an ensemble bouncing musical ideas off each other (Section 6.5.2). Second, interactivity allows an individual to feel a sense of personal agency over an experience as it is unique to them. (This latter reason depends upon the looser definition of personal agency of Section 6.4.3 which excluded the requirement of intentionality over an outcome.)

We saw in Section 6.1 Spiegel's (1988) assumption that there is a pleasure to be had simply in making music that aligned with one's 'intent'. In terms of the framework, this corresponds to a system offering 1E(live). We would assume to do so a system must be technically impressive (2T(live)) and either creating original music (2C(live)) or adapting a rendition of a stored composition (3C(prep)) such as Mathews' Radio Baton (Section 2.2.4). We questioned whether such a machine could really be enjoyable. Using the framework to draw on the lessons of other paradigms, the tentative answer is yes although with similar caveats as above: the participant must feel that this arrangement of affairs is appropriate for the context. However, we might speculate that with such a system, providing expressive agency does not compensate for poor content (C) or execution (T).

The framework also indicates the problem with Rokeby's operational clichés (Section 2.1.5). We might imagine a system that is interactive but whatever actions the participant takes, a fantastically accomplished, original and emotive outcome results. With an understanding that skill is perceived with respect to perceived agency we may see that such outcomes do not allow the participant to feel technically skilled, creative or expressive. If the opportunity to demonstrate these skills were a motivating factor in the interaction, the participant is liable to feel dissatisfied. More dangerously, however, is if the participant perceives 1 agency but later discovers their interaction was predetermined to arrive at 'good' outcomes, for example when observing another person use the system. They must then reassess the extent to which they have demonstrated skill. What they thought was their own has turned out to have been an operational cliché. There is a parallel in this situation as with the potential of a Digital Musical Instrument to dishonestly violate the concert pact (see Section 6.1.4). We might speculate from Rokeby (1998) and Schloss (2003) that overestimating skill as a result of having been 'fooled' through a misperception of agency is worse than having not perceived any skill at all. Once again, however the distribution of agency is perceived, consistency and confidence remain crucial.

6.8.1 Relationship to the Emerging Structures model

In terms of the Emerging Structures (ES) model (Section 4.2), perceived agency is closely related to a perceived capacity for action, although it also considers social and contextual norms. Perceived agency might be considered as perceived *potential* capacity defined in terms of what actions were available to the participant if they had all the relevant knowledge and skills. As such, it is crucial in determining the *value* of the participant's perceived capacity and what sense of accomplishment it warrants.

6.8.2 Perceiving agency with incremental constraint unlocking

The framework may assist our understanding of what went wrong with our attempt in the previous chapter to implement an IMS with emerging structures using incremental constraint unlocking. The IMS of the implementation provided first person agency, 1(live). However, the parameters in which this agency could be exercised were limited. We constructed the system with the goal of creating anticipation of further capacity: elements were displayed as 'locked' and written messages were presented stating when new functionality would become available. It was therefore clear that these limitations were decided upon by the designer, 3(prep). Hence 1C(live) was limited through an exercise of 3(prep) leading to some participants feeling frustrated that they were being constrained.

Although the limitations did literally limit the space of available outputs available to the participant, the framework we have presented suggests that the *distribution* of agency is more important than the actual parameters. Therefore, we may argue that had the participant not been led to attribute these to an active decision by the designer then they would not have felt as limited by these constraints. Furthermore, the constraints were controlled by a timer and so were quite clearly the result of preparatory agency. As a result, there was clearly no means by which the participant could exert influence over how the unlocking unfolded.

However, given the diversity of distributions of agency covered within this chapter, why was this one problematic? There are two points to consider. First, we observe that only some participant found the constraint unlocking an imposition. We might speculate that those who took exception were those who felt that the norms of the situation did not justify 1C(live) being limited by 3(prep). This may be the case, for example, for individuals used to using creative software on a computer and inferring those norms about this system. Second is the possibility that it was the *varying* of the parameters of 1 agency that caused the issue. The participant might have begun with a greater sense of personal agency but then had this withdrawn when the nature of constraint unlocking became apparent, resulting in a loss of perceived ownership of what they had done so far—similar to the above discussion of operational clichés.

One observation we have drawn from the framework is that there does not appear to be a universal preference for norms that maximise the parameters of agency being exercised. Rather, preference seems to be given to contexts in which the perceiver is familiar with the way in which agency is typically distributed. This suggests that limiting the user's agency is not necessarily a problem providing it does not appear to be an exercise of another's agency in violation of the user's expectations.

In Section 4.4.4 we proposed that to make an interactive music experience captivating we need perceived capacity to be continually increasing. However, in this chapter we have seen that doing so through altering the participant's perceived agency—i.e. the perceived potential capacity for action—may not be a suitable approach to do so. We may argue that this is because the participant's perception of agency in which that capacity was established is lessened, diminishing the sense of accomplishment that arises from establishing it.

To conclude, the sense of achievement gained by increasing capacity is defined by the perceived extent to which other individuals might have found it. This is determined both by the diversity of control granted to the participant and the involvement of other agents. In other words, discovering a 'hidden garden' (Arfib et al. 2003) is more rewarding if you think few others would have found it. Therefore, if the interactions available to the participant are to be constrained then this should not be done so through varying their perceived first person agency or through the unexpected involvement of another agent. In the next chapter we will apply this observation to improve on our previous design principle of incremental constraint unlocking, which will then be implemented and put to the test.

Chapter 7

Implicitly constrained interaction

In this chapter, I define an implicit constraint *as a constraint that as before limits the rate at which the participant may access the input space—but does so without altering the parameters of perceived 1 agency. I present 'mirroring', where two participants must match each other's body position, as a potential means to achieve this. This is implemented within the collaborative sound installation Impossible Alone (IA), which I describe in detail.*

The initial attempt at implementing incremental constraint unlocking in Chapter 5 suffered a number of shortcomings. The key challenge, which we identified before the study, was determining *when* the constraints should unlock. However, in the previous chapter we identified a further potential problem with limiting the parameters of perceived 1 agency (Section 6.8.2). In particular, we concluded that any limitation on the perceived potential capacity for action may be perceived as a limitation of 1(live) agency. Such limitations may be seen as an 'unjustified imposition' if they are the result of 3(prep) agency.[1]

A further issue that we identified was that the system may have been too simple. Creating IMSs for an experimental context leads to systems that are quite distinct from those that might be used outside of the lab, particularly when using quantitative methods (Marquez-Borbon et al. 2011). This is to a large extent a necessity of isolating the effect of a single aspect of a system. Evaluation methodology will be discussed in the next chapter. For now, we note simply that our model of emerging structures is designed quite closely around the notion of perceiving *complexity*. As such, it may be unreasonable to assume generalisability of results from a simplified system designed with a user study in mind.

Therefore, in this chapter I describe a system designed for public exhibition, Impossible Alone (IA) which I created with creative input from the performance artist Tiff Chan. As with the Serendiptichord, we follow an art-driven approach (Candy and Edmonds 2002), although working within the following set of requirements to overcome the issues outlined above.

7.1 Requirements

Our requirements for IA are as follows.

- The system should be constrained so as to provide emerging structures of sound and interaction. These constraints should unlock gradually in response to the participant but without an explicit goal being set.

- The participant should not perceive an alteration to the parameters of 1(live) agency as a result of the constraints. In particular, any limitations arising from the constraints should avoid the perception that they arise from 3(prep).

[1] Capacity for action was defined in Section 4.2.4. The '3(prep)' notation was defined in Section 6.4.

- The system should embody the above requirements without sacrificing its artistic merit or potential for public exhibition.

7.2 Implicit and explicit constraints

Requirements 1 and 2 together essentially demand for an approach to constraining interaction that does not alter the parameters of 1(live) agency.

We define an *explicit* constraint as a constraint that arises through a limitation *of* the parameters of 1(live) agency and an *implicit* constraint that arises through a limitation *within* the parameters of first-person agency.[2]

Thus, the implementation of constraints in Chapter 5 was an example of explicit constraints. We explicitly constrained the space of inputs available to the user which limited the parameters in which they could perceivably exercise 1 agency. As it was clear both that the input was limited and that this limitation was purposefully implemented, the constraints were likely to be perceived as an exercise of 3(prep) agency. Note, however, that perceiving agency requires perceiving *intention*. This may not be the case if the limitation seemed inevitable rather than purposeful.

So how can we create an implicit constraint? It may seem that the solution is simple not to tell the participant that new inputs will be made available. However, without expectation of further inputs we may not establish anticipation of new inputs, which was a motivation behind constraint unlocking (Section 5.1). Furthermore, some kind of feedback will be necessary to make the participant aware of this novel type of input. But by providing this feedback, and by continually providing new functionality, we have landed back where we started: the participant is aware that there are inputs that they are being denied at the whim of an unseen party.

A potential solution is to design our system so that the constraints arise due to the participant's ability to exercise fully their perceived 1 agency.

[2]There may be constraints other than implicit and explicit. For example, the participant may choose to limit the set of available inputs they use.

That is, inputs are unavailable due to the participant's ability to perform them. At the same time, the participant should be aware of the potential existence of these inputs (although not necessarily certain of what they are) and able to improve their ability during the experience in order to access them. Furthermore, we would like to avoid choosing a constraint that some participants may already be armed with skills to overcome. We cannot simply choose complex and instrumental movements.

7.3 Impossible Alone: Constraints through collaboration

7.3.1 Multiparticipant interaction

How can we constrain the exercise of 1(live) agency without imposing 3(prep) agency? Our approach will be to create a collaborative system.

Multiplayer environments are particularly popular in games (Wood et al. 2004; Mandryk 2005). Both competitive (competing against each other) and cooperative (collaborating to compete against a common goal) modes are common in games. Beyond the social aspect, an advantage of competitive multiplayer is that it allows for a range of different degrees of challenge whilst maintaining consistent parameters of 1(live) agency—in other words, the success of the players is determined only by their abilities rather than decisions of the game designer (3(prep)).

Within a collaborative IMS, requiring participants to negotiate to provide an input allows us to limit individual 1 agency ('I') but retain collective 1 agency ('we'). Through this, we propose that the limitation will no longer seem an unjustified imposition from a third party but an inevitable and necessary aspect of the interaction design.

The use of a collaborative system raises evaluation issues. For example, mutual engagement (Section 2.3.3) may lead to a more positive experience overall irrespective of the constrained interaction design. On the other hand, the need for participants to negotiate, work together and maintain awareness of each other might make their exploratory processes more conscious and so more easily probed in interview.

7.3.2 Mirroring: an implicit constraint

Mirroring is a drama exercise in which two people move together whilst continually reflecting each other's pose. The activity is designed to make each aware of their movements (Dayton 1990, p. 108). We will use the terms *mirroring*, *reflecting* and *matching* interchangeably. In each case, we mean each person positioned as if they were a reflection of the other.

Sometimes it is done with a leader but it may also be done through equally mutual influence. In the activity, each person is effectively attempting to match their own pose, but with the error introduced by each participant's perception of the other's pose and motor accuracy establishing a positive feedback loop. Thus, even without a leader movements arise naturally.

Mirroring prevents the participants from moving completely freely. At the same time, they are able easily to imagine and anticipate a greater range of inputs. Through negotiating and working together, they are free to explore the entire range of potential gestures providing them with consistent parameters of 1(live) agency (albeit collective rather than individual). Thus, requiring two participant to mirror in order to provide input to the system provides an implicit constraint. In order to access more complex movements, a rapport needs to be developed between the participants providing a pace to the interaction. Therefore progress is dependent upon skill—both social and technical. Whilst the participants remain free to establish their own goals, the challenge that they must overcome in order to achieve these goals remains clear. We propose that this constraint will lead the participants towards conducting deeper exploration whilst sustaining an awareness of the greater space, thus establishing anticipation.

The challenge provided by mirroring also allows us to introduce greater scope for 1T(live) agency.

7.3.3 Overview

Impossible Alone (IA) is a two participant IMS that uses the Mirroring constraint. It uses a skeleton-tracking system to sense the body position of both participants. Whilst the participants are not matching, there is a fairly monotonous and discordant *deviation sound*. As they tend towards a matching position, the deviation sound morphs towards a more harmonious sound. When the participants are matching, a range of more musical sounds are available through movement.

7.3.4 Process

IA was developed with creative input from Tiff Chan, a performance artist specialising in movement and dance. In a series of six workshops spread over six months, Chan and I met to test the system and discuss its future development. In some of these sessions, others were invited to participate to test the system and provide feedback. To assist offline testing, we also took recordings of input data from two participants.

Chan researches and conducts therapeutic movement classes. Throughout the process, we had slightly different viewpoints about the system, with Chan viewing it more as a movement game whilst I saw it in terms of an IMS as we have been discussing in this thesis. These differences were helpful in identifying and addressing what kind of system we were creating and how we wanted it to be perceived by participants (discussions which contributed to the arguments we have already met in this thesis).

Through this development process, two different versions of the mapping were developed. Both are described below.

7.3.5 Implementation

Joint capture

We describe a participant's body position in terms of their joints. This provides a 'stick-man' like representation, sensing of which is described as *skeleton tracking*.

IA was implemented in C++ using openFrameworks[3] with Ableton Live[4] providing sound synthesis. Sensing was done using a Microsoft Kinect[5] infrared depth camera, with skeleton tracking calculated with the NITE (PrimeSense Inc. 2011) algorithm from PrimeSense[6] running on the OpenNI (2010) framework. This provided independent measurements for the torso (including its absolute orientation), elbows, hands, knees and feet. Data for joints below the hips tended to be somewhat less reliable than those above, in terms of both missing and inaccurate data.

Joint data captured through this method is less accurate than that done using motion capture (see Section 2.2.4). The single viewpoint also makes it liable to occlusion (i.e. limbs of the participant being obscured by other parts of their body or another participant). However, it has the strong advantage of being portable, easy to set up and significantly less expensive.

The joint data it provides is noisy. Through experimentation, we determined that low latency was more important than spatial accuracy with contactless bodily interaction (i.e. without touching a physical controller). However, when considering the positions of the joints in relation to each other we found accuracy became more important at low velocities and when measuring acceleration in order to establish a sense of reactive control. In other words, when moving, we would be sensitive to the effects of noise in measuring acceleration, and when still we would become sensitive to noise within the relative position between joints. To address this, a smoothing tech-

[3] `http://www.openframeworks.cc`
[4] `http://www.ableton.com/live`
[5] `http://www.xbox.com/kinect`
[6] `http://www.primesense.com/`

nique was developed that combined a double exponential estimator (Wilson 2006; Kalekar 2004) with a weighted moving window average. This allowed us to tune a balance between sluggishness (latency on the first and second derivative) and exaggeration (oversensitive second derivative with a result as if limbs were elastic). This approach was compared to that using a Kalman Filter (see e.g. Simon 2001) and found to be superior for our purposes.

Measuring deviation

We define *deviation* as the extent to which participants are not matching. As matching was the key skill to be developed for the participants to explore the input space, deriving an algorithm that closely aligned with the participants perceptual interpretation of 'mirroring' was crucial in order to ensure their limited ability to do so was a result of their own abilities (1 agency) rather than a failure of the system (3(prep) or potentially 2(live)). To do so, considerable time was spent developing, testing and tuning an algorithm to determine deviation. We refer to a participant's joint's *opposite* as the other participant's reflected limb (e.g. for participants A and B, A's left elbow opposes B's right elbow). A joint and its opposite forms an *opposing pair* of joints.

The original proposal for IA had the two participants separated by two skeleton tracking cameras, leaving a distance of approximately four metres between them. Quite early through the workshops we found that rapport, communication and eventually mutual engagement were easier and more enjoyable to develop without the distance or equipment separating the two participants. However, this presented difficulties in capturing joints with a single camera. We wanted the participants to look at each other but also avoid occlusion. We ultimately compromised with participants stood 45 degrees apart, which was marked out with tape on the floor.

We also observed that what we perceptually considered to be matching changed based on distance. When separated by a few metres and asked to reflect each other, participants would align themselves through matching the angle of opposing joints. However, when closer together, matching was instead done through imagining a mirror was placed between them. This would often lead to different joint angles, especially with participants of different heights when, for example, raising their arms. Furthermore, the ability to maintain an accurate match dropped dramatically when joints were moving. Instead, participants (and spectators) tended to perceive similar speeds in opposing joints—with their position measured relative to the torso—as determining the extent to which they were matching.

The final algorithm conducted a joint by joint comparison drawing on a number of methods: the 3D angle created by the joint relative to its parent joint, speed and *mirror deviation*. Mirror deviation was calculated by defining an ideal reflective plane (i.e. an imaginary mirror) as the plane defined by the mean of the midpoints between opposing joints normal to the mean of the lines connecting opposing joints. The mirror deviation for a given joint was then calculated as the distance between opposing joints when one had been reflected in this plane. The final value of *deviation* was arrived at through taking a weighted combination of these methods using the L^p norm with $p = 0.43$. This value biases the combined measure towards the method that is presently most 'generous'. The overall measure was then calculated through a weighted mean over the joints. Individual joint weightings were again derived empirically through the workshops to arrive at the most perceptually satisfactory measurement. Overall, they prioritised the mirror deviation method and the hands and arms. Deviation measurements were exponentially smoothed with smoothing parameters set on a per-method basis. These weightings are shown in Table 7.1.

Mapping: Version 1

The first version was more game-like than the second (described below). As participants moved without deviating, a 'charge' built up, which was sonified as a deep electrical buzzing growing in volume and intensity as the charge increased. If a charge was built up and the participants then deviated, it

Table 7.1: Weightings used to calculate deviation.

Method	Weight	Joint	Weight
3D angle deviation	0.43	Torso	0.38
Speed deviation	0.1	Elbow	0.23
Mirror deviation	0.47	Hand	0.23
		Foot	0.015

would be 'dissipated': the charge would rapidly decrease and the system emit a deep crack of thunder would resound through the room, proportional in intensity to the level of charge.

The soundscape was divided into a sequence of levels which were activated when the charge reached a given threshold. At each level, different tracks of music—mostly loops ranging from a few seconds to a few minutes—would come in as well as new inputs recognised, providing control over the volume and effect parameters of some of the tracks. As the participants progressed through different levels, the requirements to build up charge would become more stringent, eventually to the stage where continuous movement above a certain speed was necessary to avoid dissipation.

The charge and level system was motivated by the desire to form musical structures of tension and release within the system. However, advancing through the levels became an overriding goal that detracted participants' engagement with the more musical functionality. Although the participants remained free to move as they liked, they built up charge providing it was with enough speed, and it was this build up that became the primary goal rather than exploring the mapping. Different ways of 'gaming' the system could be found such as repeating simple arm movements to rapidly advance. The music served more to indicate the achievement of the goals rather than the focus of attention.

Furthermore, the levels within the system effectively served as an explicit constraint. Although this early version was not formally evaluated as was the system in Chapter 5, the issues of perceived agency did not arise. However, this is likely because the system was perceived much more akin to a game than a tool for personal expression leading to different norms of agency being inferred by the participants.

This first version did not establish an implicit constraint as intended. However, its development was informative in identifying how introducing game aspects—in particular a level structure—can alter the attitude and exploratory approach of participants.

Mapping: Version 2

In order to retract some of the game-like elements and create a more music-oriented system, the charge and levels were dropped in the second version.

The development of this mapping was more strongly influenced by the effect of the matching constraint on participants' movements. In the workshops, we found that this caused a tendency towards certain types of movements, in particular in the order in which limb movements were tried. Participants tended to commence with arms with torso and leg movements following later.

We began with the idea of interpreting abstract qualities of the movement of the participants, working with Laban Movement Analysis (Davies 2001) and affect recognition (e.g. Bernhardt and Robinson 2007; Kapur et al. 2005). However, over the course of the workshops, we moved towards more reactive and easily understood mappings.

This mapping was developed through a number of iterations. The result was as follows. The deviation sound fades out when participants are matching and moving. If participants deviate, or stay still for more than a few moments, the musical sounds (described below) fade out and the deviation sound returns. As the mapping is only effective when matching, we may define it in terms of one participant.[7] Relative to a participant, we define

[7]As we saw, there is some leeway in our deviation measurement. In practice, we reflect the second participant in the reflective plane and use the mean position of opposing joints within our mapping.

three planes: vertical (left/right and up/down), horizontal (front/back and left/right) and sagittal (front/back and up/down) (Davies 2001).

- The hands are mapped to an 8 bar looped string section with the position of each hand in the vertical plane controlling timbral effects (low-pass filter frequency and resonance, FM drive and harmonic density). The mean speed of both hands control the volume of the strings. Through experimenting, we found that using the mean of the current velocity and a decaying peak velocity provided a satisfying 'conductor-style' means of controlling the level.

- The contraction of the arms (i.e. the distance of the hands from any part of the torso) controls the level of a rougher pad sound, with the level decreasing as the arms contracted.

- 'Stabbing' the hands below the height of the hips creates a deep percussive hit. The intensity of the hit could be controlled by 'accumulating' energy through fast arm movement before the stab took place. Note that this is not quite the same as the velocity of the arm at the moment of the hit determining intensity—this discrepancy from what might be expected confused some participants in the evaluation. If the hands are kept below the hip line after a stab then moving them within the horizontal plane gives control over effects applied to the decay of the hit sound.

- A generative two bar drum loop can be triggered by simultaneously passing the hands and elbows through the horizontal plane at the height of the shoulders, with the algorithm's two input parameters controlled by the mean velocities of these joints in the up/down axis and the horizontal plane. Each arm controls a different drum (and each participant had different drums providing four in total, although two would typically be triggered concurrently due to matching). The drum sounds pass through an effect which controlled by the mean up/down speed and mean elbow angles. Triggering of the drum loop is quantised to a semiquaver, which is the shortest note duration within the loop. Extra loops may be created and triggered simultaneously by moving the arms out of and back into the trigger zone. Through this, complex rhythms can be created through moving the arms as if imitating a bird flapping its wings.

- Clapping the hands triggers a short generative sequence of bell sounds. Inaccuracies in the skeleton tracking made it difficult to determine when an actual clap occurs so a threshold of hand proximity was used as a trigger. As the hands move towards the threshold, a 'tension' sound of a steady pitch can be heard, with its volume linked to the closeness of the hand distance to the threshold. After a clap is detected, the closeness of the hands controls the phase on a (non-oscillating) phaser.

7.3.6 Exhibition

Version 1 of the system was shown at the Secret Garden Party music festival in July 2011 with the Guerilla Science group (Figure 7.1), where it received a favourable review from New Scientist's CultureLab (Else 2010). See Section 1.5.2 for a full listing of exhibitions.

7.4 Conclusion

In this chapter we have established *implicit* constraints as a means to implement incremental constraint unlocking without disrupting the user's perception of 1(live) agency. We have realised this concept through the mirroring constraint. As well as being an implicit constraint, mirroring brings other aspects to participation. From my experiences so far, I can report that it establishes a strong sense of mutual engagement, creates a situation in which both participants provoke each other into experimenting with unfamiliar movements. It has a strong disinhibitory effect. However, this is potentially illustrative of a shift from singular 1 agency ('I') to collective

Figure 7.1: Two participants interact with Impossible Alone in front of a crowd at the Secret Garden Party, 2011.

(‘we’).

Collaborating with an artist from a different background highlighted where assumptions were being made. In particular, it made us address potential trade-offs between music-oriented and game-oriented aspects. My approach with IA has been to create from the ground up a system that embodies an implicit constraint, rather than attempting to ‘bolt on’ constraint unlocking to an existing idea. I have argued that this is the fairest way to assess the real-world potential of our approach to creating interactive music. However, conducting a rigorous evaluation of such a system—and the principles underlying it presents a challenge, one that we will tackle in the next chapter.

Chapter 8

Evaluation of Impossible Alone

In this chapter I present an evaluation of the previous chapter's implementation of an implicitly constrained interaction design, Impossible Alone (IA). The outcomes of this evaluation have informed the theoretical work in earlier chapters. Before doing so, a number of shortcomings of the quantitative approach to evaluation used in Chapter 5 are discussed, with the conclusion that a rigorous qualitative method is most appropriate. I opt for Discourse Analysis (DA), review a number of different approaches to this method and opt for an adaptation of an approach taken by Stowell (2010). I report on a mixed set of experiences and present the key findings of the study in line with the earlier work of this thesis, exploration and perceiving agency, as well as some interesting findings regarding the rights and obligations perceived by the participants. Finally, I offer an informal evaluation of our methodology.

In Section 5.6 we identified a number of shortcomings with the constraint unlocking study. The approach to constraint unlocking limited perceived agency (see Section 7.2). Furthermore, the constraint implementation did not sufficiently reduce the complexity of our interface (see Section 5.6). However, there were also a number of issues with our methodology. Before describing the evaluation of Impossible Alone (IA), we therefore will critically analyse our earlier methodology to best consider how to proceed.

8.1 Methodology

Engineering for pleasure requires different approaches than when the goal is utility (Gaver et al. 2004). In this section we will outline a number of shortcomings that can arise when quantitative approaches are used to evaluate creative or aesthetically-driven systems. I will then propose that a qualitative approach is more suitable for our needs.

8.1.1 Problems with quantitative evaluation

Quantitative evaluation has a key advantage in that it allows the researcher's subjectivity to be detached from data analysis. Whilst few argue for art to be evaluated through quantitative methods (Höök et al. 2003), it is common when evaluating technology used within creative practice to 'factor out' the aesthetic components and just consider utilitarian aspects such as usability, learnability and efficiency (e.g. Wanderley and Orio 2002; Jordà 2004). However, I will argue here that quantitative methods are unsuitable for the evaluation of a design principle for use in a creative work. There are four key issues which we will discuss in turn.

1. Evaluating design principles based on the mean response of an audience.
2. Assuming audience responses are independent.
3. Neglecting the situatedness of interaction.
4. The need for a creative work in the evaluation.

1. Evaluating design principles based on the mean response of an audience

The mean response of an audience does not necessarily determine the success of an artwork (see, e.g., Wellcome Trust 2010). A work may not be made with the goal of eliciting a single homogenous experience (Morrison et al. 2007). The average experience reported by statistical methods is not necessarily representative of a single experience (Gaver et al. 2004). As Höök et al. (2003) points out, it is the anomalies, complexities and individual peculiarities that are of interest.

More helpful is an analysis of *why* a work elicited a particular experience. This is not in conflict with our goal of establishing generalisable approaches that may be applied to create specific effects (Section 1.4). Rather, that we may better achieve it by going beyond statistical aggregation and considering individual experiences in more detail.

2. Assuming audience responses are independent

Quantitative studies typically assume independence for each datapoint. This is usually a requirement when testing for significance (Kruskal 1988). But appraisals of art work often are not independent. Our response is influenced by our preconceptions and expectations (Leder et al. 2004), many of which may be formed by responses of others, media coverage, critical reviews. The behaviour of others within a gallery may influence our opinion as may knowledge of the artist, as we saw in Section 6.2. It is not uncommon to attend galleries with friends and there may be mutual influence within a group. Not only are these important factors to consider, but they likely interact with each other too. For example, a friend's recommendation may be less influential if we have already heard about a work from the press.

Therefore it may not be reasonable to assume that the success of an artwork is determined by its ability to be successful with many independent viewers. If the responses of individuals are mutually dependent, then it is not valid to assume it is a population (in the statistical sense) that may be independently sampled. This calls into question the validity of generalising from the lab into the wider world.

3. Neglecting the situatedness of interaction

Throughout this thesis we have considered the means by which an aesthetic experience is established through how we form an understanding of a system. Suchman (1987) argues that the actions we take to explore a system, and thus the means by which we form an understanding of it, are *situated*, meaning that context and environment are strong determiners both of behaviour and the interpretation of its consequences. In addition, how individuals respond aesthetically has been shown to be dependent upon the environment of reception (Leder et al. 2004).

It is not always possible to test things in the field, and lab-based research remains a vital tool. But this change in context limits the extent to which we may generalise conclusive results of *whether* something was the case. However, with a richer understanding of *how* experiences are formed and systems explored we can more easily theorise the effect of the situation.

4. The need for creative work in the evaluation

In order to evaluate engineering principles that are developed for creative applications, creative works need to be involved.

When evaluating a design principle through a creative work, the researcher's artistic ideas become entwined within the evaluation (for it usually is the researcher's own work that is being evaluated). Even DMIs embody musical ideas (Jordà 2005). As artistic goals may limit the validity of the results, they may well remain undisclosed. But art is typically created to serve some kind of greater artistic goal. A particular design principle may or may not contribute to that goal. I would argue that those who may be interested in applying the outcomes of such research are likely to be interested in the design principle in terms of how it assisted the artist in achieving these goals. Evaluating it against a general 'positive user experience' metric as we did in Chapter 5 may make the results easier to interpret and compare. However, it obscures the extent to which the design principle contributed towards any artistic goals.

Furthermore, evaluation of an attribute of a system under the standard hypothesis testing paradigm requires a control condition to be established that ideally varies only in terms of this attribute. Therefore, the researcher must convince their audience that any effects recorded are due to this attribute rather than other design decisions. The simplest means to do so is to disable a particular functionality of a system that embodies the design principle. However, the earlier on during the design of a system that a design principle is adopted, the more sympathetic the design as a whole will be to that principle. Decisions taken during design depend upon each other and attempting to create a system that works both with and without that principle may lead to a compromised system that does not work in either case.

We saw a potential example of this issue in Chapter 5 with our systems that were considered too complex by some participants but were perhaps still too simple for incremental constraint unlocking to be effective (Section 5.6). Constraint unlocking is a means to continually introduce complexity. When a commitment has been made to it, there is more opportunity—arguably a need—to introduce complexity into the design.

Quantitative testing is designed to evaluate the effect of a treatment on a population. But the population that design principles are applied to is *interactive artworks*, not the individuals who experience a single work. If we were truly to evaluate the principle quantitatively, we would need to sample from the population of 'potential artworks', recruit a team of artists and somehow convince half of them to implement the principle in their work without biasing the rest of their input. It is an absurd idea. Otherwise, however, the only generalising power provided by our statistical tests concern the effect on the audience of *this* particular work.

Whilst quantitative methods remain important within NIME research for evaluating traditional HCI issues such as usability, efficiency and learnability, as we start to consider the interactions between these issues and aesthetic outcomes, more subjective and individualistic influences become involved. When we are evaluating the experience of interacting with an IMS, a large number of assumptions are necessary to maintain the validity or generalisability of any findings. As we have seen, many of these assumptions may not be valid.

As a result I opted to conduct a qualitative evaluation of Impossible Alone (IA).

8.1.2 Qualitative approaches

Whilst quantitative research focuses on finding generalisations across populations, qualitative approaches often privilege the individual's point of view, seeking rich descriptions that capture the constraints of everyday life with all of its interacting aspects (Denzin and Lincoln 2005).

It follows an open approach where the potential outcomes of the research are not predefined before it is undertaken, making unexpected findings more probable. Perhaps most importantly from our point of view, it facilitates interpretation of *how* data arises from a specific individual's experience.

I argued above that when design principles are applied to artistic works, applications are likely to differ both in implementation and in intention. Therefore results derived from a single work may lack the generalisability often associated with quantitative methods. However with a qualitative approach, whilst the specific outcomes may not offer generalisability, we provide descriptive details and interpretations of case-studies allowing us to understand *why* events turned out as they did. This allows others to adapt the lessons for their own situation. It is through considering the many facets of an individual's experience that we may understand the role their specific opinions, idiosyncrasies and experiences played in its construction. We may also interpret the extent to which specific details of our implementation played a part. As a result, qualitative approaches allow us, if not to avoid, at least to recognise non-generalisable aspects of our result. For this reason, we might argue that in some contexts they may provide *more* scope for valid generalisation.

Limitations

A key limitation of qualitative research is that it requires a significant amount of subjective interpretation by the analyst (i.e. the researcher). As a result, there is potential for the preconceptions and biases of the researcher to distort or misrepresent the data. There are, however, procedures to help limit this possibility. As mentioned in Section 2.5.5, there are diverse levels of rigour to any approach and procedural documentation is essential to allow a reader to assess the validity of the analysis. Our procedure is documented below in Section 8.3.

A consequence of the above is that further analysis is required of the reader to ascertain that the interpretation provided is a fair interpretation of the data.

8.2 Discourse analysis

Discourse Analysis (DA) describes a range of techniques for detailed analysis of text. There is no single fixed method (Banister et al. 1994). However, the term DA is used within the social sciences to demand a depth of analysis beyond simply reading and interpreting the transcript (Antaki et al. 2003).

DA was popularised by Foucault (1969), although he did not define a specific set of methods and a number of differing approaches have formed (Peräkylä 2005). Common to these approaches is the application of a systematic analytical method to uncover not only the surface meaning of the text, but how language is used to construct a common reality (Banister et al. 1994). As a result, discourse analysts often talk about the 'work' a text is doing (e.g. Stowell et al. 2008; Holstein and Gubrium 2005) or the 'machinery' of language (e.g. Banister et al. 1994; Silverman 1998).

Key to DA is the viewpoint that there is not a unified concrete reality hiding inside an individual which may be 'accessed' through an interview as if they were a noisy measuring device. Rather, talking and interacting are a fundamental part of *forming* subjective reality (Peräkylä 2005) and understanding our experiences. It is through discourse that we construct our reality and bring meaning to our experiences (Holstein and Gubrium 2005).

Our approach will be to rigorously understand the reality that is constructed through the discourse of a post-experiential interview. This does not imply that we are uncovering 'what really happened'. As the experience is put into words meaning is imposed upon it. It is rationalised and no longer an experience but a reconstructed experience from a memory. However, it does still provide insight into how different aspects of a work elicited different responses from its participants, as well as how a participant relates to and draws upon previous experiences when forming an understanding of this experience (Stowell 2010).

8.2.1 Analytical approaches

One of the strengths of DA is its adaptability to differing needs. However, this adaptability means that there is not a unified systematic approach that may be followed (Inskip 2010). While rough guidelines have been provided (e.g. Banister et al. 1994; Baker 2004), these tend to be provided more as examples to aid those intending to apply the technique in developing an *analytic mentality* rather than a set of stipulations to identify 'bad' analysis (Potter 2004).

Being open to a variety of methods for analysing text does not mean that 'anything goes'. In particular, analysis needs be *systematic* and *explicit*, going beyond simply summarising, pulling quotes or forming an opinion based around the data (Antaki et al. 2003).

Unlike, for example, grounded theory (Section 2.5.5), DA does not require the analyst to 'suspend' their experience, knowledge and prejudice when analysing the data (Inskip 2010). But the risk remains that, either intentionally or accidentally, the analysis technique is applied simply to construct evidence of the preconceived opinions of the researcher. Awareness of this risk becomes particularly relevant when we wish to use DA methods in fields that have traditionally adopted more quantitative approaches. Therefore those applying DA within an engineering context, such as Stowell (2010) (whose method ours is based upon) and Inskip (2010), have adopted highly systematic and thorough methods that are explicitly documented, demonstrating not only how, but also the extent to which the discourse was analysed.

Although discourse analysis techniques have been applied directly to the interaction between humans and machines (e.g. Suchman 1987), we are interested in understanding not just how someone interacts but how that interaction makes them feel, what they think about it and what experience they take away with them. In Section 2.5.5 we established that that think aloud methods are not appropriate. But trying to determine this entirely from what participants *do* gives an incomplete picture. Therefore we have chosen to analyse the experience after it has happened through a reflective interview.

Interviews are common within qualitative research as they allow access to past events and experiences (Peräkylä 2005). However, it is important not to overstep this and assume that we are somehow analysing the experience itself. The representation of the experience is being constructed in the interview (Baker 2004) and it is this discursive representation that we are analysing. However, by producing a rigorous interpretation of our participants' discourse we may make *inferences* of their prior subjective experience and how it was formed through their interactions with a system.

To summarise the above, the key requirements we require of our analytical technique are as follows.

1. It involves a depth of analysis that goes beyond the surface meaning of the text.

2. It is applied in an explicit and systematic manner that limits the potential for distortion or omission of parts of the data.

3. It is open to arriving at unexpected conclusions when they are supported by the data.

4. The validity of the above may be demonstrated to a reader.

8.2.2 Stowell's method

Stowell et al. (2009) build on the approach described by Banister et al. (1994, ch. 6) to present a method of analysis that that involves a rigorous and systematic procedure to extract predicate relations from the discourse. The ultimate object of study is not the discourse but it is through understanding the discourse that we hope to understand the events that caused it. In this sense, we might consider it *applied DA* in that the theories and methods developed within DA are utilised but with different objectives and outcomes. The method is as follows.

1. *Transcription.* The interviews are transcribed in a fashion that includes all speech events (repetitions, fragments, pauses, simultaneous talking from different individuals). These transcripts should be made available to the reader to allow for analytical transparency.

2. *Free association.* The analyst reads the transcripts and notes down surface impressions.

3. *Itemisation of transcribed data.* All pronouns are resolved using the participant's own language as far as possible. The *items* and *actors* (which includes objects that act with agency or sentience) that appear in the transcript are identified. For each of these, a description is extracted from the speech, again using the participant's own terminology as far as possible.

4. *Reconstruction of the described world.* The participant's described 'world' is reconstructed into a conceptual map, such as a network of relations. The participant's language is retained throughout and so

this world is primarily a methodical reorganisation of the participant's language through close reading of the text.

5. *Examining context.* The analyst may compare between different discourses and draw upon their knowledge of the world beyond that described by the discourse.

A key strength of this method is that it addresses the *entirety* of the data. Every utterance is systematically included as a predicate, item or description, or explicitly excluded as a conversational mechanism. The data is analysed in a strict and methodical fashion that highlights where subjective decisions need to be taken, allowing the analyst to avoid unintentional bias.

The method is particularly suited to analyse retrospectively an experience that could not be addressed directly, as is the case in our context. As such, an amount of data is disregarded: the discourse is 'flattened' in time to produce a static conceptual map. Many of the mechanics of conversation are considered less relevant in uncovering this aspect and disregarded.

Flexibility

We argued in Section 8.2.1 that an advantage of DA is its flexibility, yet also of the need to be systematic. These need not be contradictory requirements. With the above analysis, we have outlined our means of ensuring that the entirety of the data is systematically covered. By making this process explicit, we are able to identify where data is being disregarded. In particular, where it becomes apparent that our 'flattening' of the data is disregarding potentially valuable information, the analyst is free to include extra details within the analysis.

Subjectivity and transparency of analysis

DA is not a magic bullet that turns discourse into objective data. However methodical, analysis is done by an analyst. The strength of Stowell's method is not in removing this subjectivity but in making it explicit, making the analyst conscious of it and ensuring it is applied to the entirety of the data rather than to parts that might initially look 'interesting' (i.e. useful!).

As alluded to above, with such a diversity of approaches it is important that, as well as being systematic, analytical methods are documented explicitly (van Dijk 1990). Where the term 'DA' is used without any description whatsoever of an analysis (e.g. Deweppe et al. 2009), as readers we cannot determine the extent to which a structured method has been used. As well as documenting our procedure, in line with Stowell (2010), full transcripts of the interviews used in our analysis are available online (see Section 8.4).

8.3 Method

Our evaluation was conducted using IA, described in Chapter 7. We are particularly interested in identifying the effects of the matching constraint. As discussed in Section 8.1.1, we cannot assume that simply 'removing' this constraint will leave us with a plausible system. The mapping has been constructed *around* this constraint (Section 7.3.5). However, in order to allow participants to better understand the effect of the constraint, they were also asked to use the system without the constraint.

The orders were not randomised (so as to not infringe upon the freshness of the 'actual' experience). Furthermore, at the point when the participants used system without this constraint they had already been interviewed and asked to consider its effects. This second session was conducted less to be an experience for them to report on and more as means to allow them to further develop their understanding of the effect the constraint had on their first experience.

8.3.1 Pilot studies

Three pilot studies were conducted which uncovered a range of practical issues, primarily regarding the system design. A short extract of an interview was transcribed and analysed in order to identify any prominent issues within the analytical procedure.

8.3.2 Participants

Six pairs of participants took part in the study.

In order to maximise the participants' potential to establish a rapport within the study, participants were chosen based on their prior skills as far as possible. However, the pilot studies had indicated that the deviation algorithm deteriorated as the difference in height between the participants increased. Therefore, maintaining similar heights was given preference to matching skill when pairing participants.

Participants were recruited through sending recruitment adverts to mailing lists. Participants were invited if responding to provide their height as well as self-assessed ratings on their music and dance experiences on a five point scale ranging from *None* to *Professional*.

One respondent requested to take part with her friend, which was accommodated. Two of the other pairs were aware of each other from working in the same department, but unacquainted. The other three pairs did not know each other.

Participants were each paid £10 for their participation, which ranged from approximately 25 to 75 minutes.

8.3.3 Conditions

Although we did not consider the effects of prior knowledge with the Emerging Structures (ES) model, we saw both in the constraint unlocking user study (Chapter 5) and then with the framework of perceived agency a participant's expectations of a performance or interaction can play a crucial role in their response. Therefore, to investigate this effect, half of the pairs of participants were provided with an overview of the available mappings, half of them were not. Pairs of participants were randomly assigned to either group in equal number.

We emphasise that this was not an experiment. Our interest is less in generalising outcomes and more to understand the processes that led to them.

8.3.4 Data collection

Each session within the study was video recorded. The sound output from the system was recorded separately and a portable sound recorder was used to record the interviews (in addition to the video recording in order to provide a backup).

Whilst participants were interacting with the system, observations were recorded. However, in order to minimise any inhibitory effects, this was done primarily through looking at a video display of the participants on a laptop rather than at the participants directly.

The consent form provided to each of the participants included options to allow them to opt out of the normal anonymity requirements of the data being recorded in the study. It was made clear that these answers could be changed after the study. Approximately half of the participants chose to do so.

8.3.5 Procedure

The following procedure was followed.

1. Participants were welcomed. An overview of the study was provided and the participants were asked to read, fill out and sign the consent form. The overview did not include details of Impossible Alone beyond describing it as a musical system that they would use.

2. An overview of the system was provided that covered the following.

   (a) The system responds to their movements with sound.

   (b) To hear the sounds, the participants had to match each other's bodily position as if there were a mirror separating them, and could not remain still for more than a few moments.

(c) There would be a 'dull feedback sound' to inform them when they were deviating or staying too still.

(d) The system does not respond properly if their limbs are not visible to the depth camera or if they move beyond the range of the depth camera (marked by tape on the floor).

(e) If applicable, participants were given an overview of each input of the mapping of the system. Otherwise, participants were told that they would not have the mapping explained to them.

(f) The participants were free to use the system for as long as they like and should inform the investigator when they would like to stop.

3. The participants were then asked to calibrate with the system[1] and the system was then switched on.

4. After the participants had indicated they would like to stop, a group interview took place. During this, the investigator avoided as far as possible to influence what was said beyond probing certain issues and steering the discussion towards topics of interest. In particular, care was taken to use terminology introduced by the participants wherever possible and avoid potentially 'loaded' terms (deMarrais 2004). Note that such an approach is an ideal. In practice, it is inevitable that an interviewer's influence will extend to some extent beyond the general topics that are discussed.

5. The participants were invited to use the system again but informed that this time they would not need to follow the matching requirement.

6. Steps 3 and 4 were repeated.

7. The participants were thanked for their time and paid.

[1]The OpenNI (2010) software used in this study required participants to calibrate by holding a specific pose for a few seconds before they would be tracked.

8.3.6 Analysis

Our method of analysis builds upon that of Stowell (2010) (described in Section 8.2.2) and is as follows.

1. The discourse is transcribed. Note that the extent of detail used within some transcripts is not always practical when large amounts of discourse is involved due to time constraints (Inskip 2010). Our transcriptions include all details such as fragments, interruptions, overlapping speech, pauses. However, it does not include the timings of pauses, changes in the speed or pitch of talking, or emphases except where necessary in order to understanding the surface meaning of the text.

2. The entire discourse is read through, with initial thoughts and points of interest recorded.

3. The text is broken down into utterances that are entered into a spreadsheet. During this process, potential item categories are noted.

4. The text is itemised: the items and actors used are grouped into categories across utterances. The description used is saved with each category.

5. Predicates relating categories are extracted from each utterance (more than one predicate may follow from a single utterance). Predicate descriptions are constructed from the text of the utterance as far as possible.

    - Where utterances relate more than two items, two items are chosen as subject and object by the analyst with other items kept within the description. This decision is made to preserve the original meaning of the predicate as far as possible.
    - On other occasions there may only be a subject and no object.
    - Where participants have constructed a sentence together then the predicates are created for both speakers.

- Where an utterance is a concurrence of a prior utterance from another person then it is disregarded unless it actively goes beyond normal conversational mechanics, in which case the predicate to which it concurred would be created for both speakers. For example, 'yeah' would be disregarded; 'yeah, it was' would possibly be disregarded depending on the analyst's interpretation; 'yeah it was, wasn't it' would not be disregarded.
- Note that as the predicate descriptions were simply extracted from the participants' words it is not necessary to interpret a logical meaning or even to fully understand what was said.

6. The resulting predicates are used to construct a conceptual map for each speaker where nodes represent items and are labelled by all descriptions used for that item by that speaker, and arcs represent relations and are labelled with the descriptions.

   These maps are often quite complex. Part of this process involves organising the map into clusters to keep it as simple as possible.

7. From the map, contradictions and ambiguities may be identified.

8. A description of the conceptual map is recreated from *all* of the predicates on it in the language of the participant. Note that such a summary is not the 'result' of the analysis. Rather, it is an outcome that is useful in providing a non-temporal description of the participant's contribution to the interview. It also assists a reader their verification of the analytical process and conclusions that are drawn.

   A motivation for using this technique to recreate the discourse is that it preserves and aids the identification of ambiguity and contradictions in what was uttered as they will tend to have been clustered together during the construction of the map.

9. The above procedure is repeated for both interviews. A summary of the key findings from this pair of participants is then written, drawing

on all of the above analytical material as well as the analyst's own knowledge and noted observations from the study.

Note that as well as analysing what was said, following the above procedure further serves as a means for the analyst to become fully immersed in the data.

As mentioned in Section 8.2.2, whilst our method is systematic to ensure we cover the entirety of the data, on occasions it may become apparent that it is disregarding important information. In these cases, the analyst remains free to include additional material as a part of the analysis.

8.4 Results

We present here the findings from each analysis (see Section 8.3.6, step 9). Details of observations and the reconstructed summaries may be found in Appendix C. Full transcripts of the interviews are available online `http://timmb.com/phd/transcripts`.

In this study participants were initially stood side by side. Based on this, they are described within each study as L and R referring to the participant on the left or right respectively. All skill ratings are those provided by the participant in their email and in the range 0–4 inclusive.

Note that the results of the analyses from this chapter led to further refinements Chapters 4 and 6. As such, the approach taken by the interviewer as well as the language and findings may seem undeveloped in light of the material we have presented so far. However, I have opted to keep our presentation of the results in line with the original analysis so as to allow the reader to understand the way in which our earlier theoretical content is supported by this data.

8.4.1 Participant pair 1

These participants had the mapping explained to them before they began. The first sessions lasted five minutes and the second nine minutes.

Participant	Gender	Musical skill	Dance skill
L	F	0	1
R	F	2	1

Findings

The participants did not seem to develop a strong rapport with each other. R was more dominant while they were using it and both sessions were ended with R's lead.

There was quite a difference here between the participants in that L seemed to feel she was supposed to be adapting to the music and fitting it. R on the other hand seemed to have embraced the idea that the music should be following her. However, she seemed to have set herself the goal of understanding how the mapping works and was frustrated not to be able to do so. Whilst she reported briefly losing herself in some of the collaborative moments, she implied that the system was obliged to behave in a consistent way, to allow her to explore its affordances and to use it to express her own creativity. She reported deciding to stop using the system when she could not go any further with this. In the second interview, R expanded on this stating that in the first session the participants were trying to *'crack'* the system. But in the second, not only did she report that they went 'deeper', but also that they *'happened upon'* something new.

The new thing that they discovered was what it sounded like when they were very still. This seems slightly counterintuitive - a key aspect of matching is that it should slow down the participants, so we would expect sounds from very small amounts of movement to have arisen in the first session. But, perhaps matching gave them pressure to keep up a constant movement. Either way, discovering something that she had not expected to find seemed to provoke a more positive experience than those aspects that had been explained. It is tempting to conclude that explaining the mapping establishes an implicit task of trying everything out and then making something of it whereas leaving things to be serendipitous discovery does not.

In this sense, the constraints seemed to have failed in their objective. The participants were not slowed down but were (for L at least) limited from repeating earlier discoveries.

Overall, L did not speak too much in the interview, but her general response seemed to be dissatisfaction with what she heard. She felt the music should be inspiring her to move (rather than the other around), and its failure to do so became more of an issue when she did not have a partner to follow.

8.4.2 Participant pair 2

Participant	Gender	Musical skill	Dance skill
L	M	3	1
R	M	2	1

The participants had the mapping explained to them before they began. The first session lasted five minutes and the second four minutes.

Findings

The participants seemed to have a good rapport.

The participants were both very positive about the experience after the first session. A fair bit of the discussion revolved around how to classify the system. L made a number of connections with music production but was reluctant to consider it more than a way of generating ideas. Both participants, however, felt it was closer to a game than anything else.

In spite of this, the participants reported much preferring the second session. This time, they report a much more musical experience. Both participants report this session as being more about the sound than the movement, indicating transparency developing. These differences are shown on closer analysis of the text, where we see a shift around line 29 where intention becomes expressed in terms of sound output rather than intended movements or triggers.

We also find a difference drawn out here between externally motivated (*'you had to')* and internally motivated (*'you chose to'*) cooperation suggesting a loss of perceived personal agency when matching. This would suggest that the constraint of matching was not perceived as 'implicit' as intended.

On a number of occasions, the interviewer attempted to draw out distinctions between different types of exploratory approaches (broad and deep) but the participants did not respond to this idea.

8.4.3 Participant pair 3

Participant	Gender	Musical skill	Dance skill
L	M	4	3
R	M	4	2

The participants did not have the mapping explained to them before they began. Both sessions lasted 19 minutes.

Findings

The participants had a good rapport but did not seem to make much eye contact. As a result they had observable difficulty in communicating intention.

They had a lot to say in the interview. R mentioned that he had done similar sorts of mirroring exercises before at drama school. His idea of mirroring was more akin with how we had originally intended it—making eye contact and starting off with slow movements. However, he was frustrated that he could not make eye contact and that they had started off right from the beginning at a quick pace. He felt that as a result of this they got tired too soon and lost momentum. This indicates that the matching constraint may not have worked entirely as hoped.

Both participants discussed different possible uses of the system and make clear that different scenarios would impose different obligations on the system. For example, in performance the mapping would need to be known in its entirety by the performer. This given, they seemed quite unpeturbed by perceived inconsistencies in the mapping. They mentioned that the drum that is hit below the hips does not seem to respond to velocity as might be expected but that this adds to the *'glitch element'* which makes it OK. This seemed to demonstrate that perceived obligations of the system can make the difference between an unexpected behaviour being seen as frustrating, amusing, thought-provoking, requiring further investigation and so on.

R gave an interesting description of how one goes about exploring the system. *'Anchors'* are found. From these familiar points, new unknown territory is explored but when there is a lull, these are returned to. Both participants described their exploration as beginning with *'randomly exploring'* (*'just go* *mental'*) which is followed by look at more *'technical'* aspects. In this sense it seems like we can distinguish between task-driven usage (though not as task-driven as traditional HCI tasks) where either a particular outcome is expected (or being tested) and non-task-driven usage, where the input space is more randomly explored until something is experienced that the participant would like to hear again. This use of repetition was also mentioned by L as a means of testing hypotheses about the mapping and their role within it.

Overall, the participants reported preferring the non-matching version but also said that there was a strong ordering effect (R: *'I'm glad I did the first one as I think it made me appreciate it more'*, L: *'it would be confusing as a first experience'*) although they were very positive about both systems. They related the system a fair bit to experimental musical instruments (the Theremin and a number of times the Reactable). In these contexts, mirroring seemed to restrictive, primarily as both participants had to create the same sound.

8.4.4 Participant pair 4

Participant	Gender	Musical skill	Dance skill
L	M	2	3
R	F	1	0

The participants did not have the mapping explained to them beforehand. The first session lasted 16 minutes and the second 6 minutes. This was the only pair where the participants were of a different gender.

Findings

They quickly established a good rapport. These participants really pushed the system to the limits of its abilities - particularly the skeleton tracking capabilities. They were quick to try lying on the floor. At one point R stood on his head which the system struggled with (as the investigator, I also intervened at this point to ask him to stop as I was nervous that it went beyond my risk assessment). A number of occasions I also had to ask one of them to recalibrate the skeleton tracking as it was lost from them through, e.g., lying on the floor.

The participants both seemed to respond more positively to the mirroring session, with L generally being more positive than R. In particular, L described the shared experience as being uninhibiting and intimate but also safe and playful. When describing how they went about exploring, they gave a similar description as in study 3: finding things that *'worked'*, then either trying the same thing in a different context (development) or trying something totally different (broad exploration) and then returning to something they knew worked.

The participants were quite explicit in saying that they stopped because felt they had exhausted the possibilities of the system, which they attribute to not enough musical control: *'it felt too much like you were cutting or fading into a piece of music that was already playing'*. Although R mentions that they were expecting *'a controller'*, the participants generally do not describe the system in terms of being an instrument (L: *'a dynamic movement matching system'*, R: *'a responsive audio system with forced collaboration'*). They do arrive at a sense of being responsible for controlling the sound: (R: *'it is always following your moves. I found sometimes you want to react to the system as opposed to it reacting to you but you will always be chasing your tail if you try this'*). There is, however, an implication that they have the right to continually be able to develop an understanding of the system through its audio feedback (R: *'I think the feedback was a bit "blehh". I did not feel like I really worked out what was going on. I did not need to know everything but the amount of movement and exploration we did justified more.'*). R emphasises that the possibilities of the system are closely tied into the space and that he did not particularly like this space.

One interesting observation of this pair is that whereas the other participants spoke of being constrained as something the system had imposed upon them due to its limitations, this pair expressed surprise that they themselves had been restricted, for example in not leaving the marked boundary throughout the study. Their restricted movements were something *'the system led you to do'*.

8.4.5 Participant pair 5

The participants had the mapping explained to them before they began. These participants were both older than the others, being middle aged. They were also the only pair of participants that were good friends. The first session lasted 20 minutes and was ended by the investigator. The second session lasted 10 minutes.

Participant	Gender	Musical skill	Dance skill
L	F	1	1
R	F	-[a]	-[a]

[a] L invited R to attend this study as a friend. As a result, no self-assessment data was provided for R.

Findings

The participants established before they began that R would lead and L would copy. They initially began by doing familiar dance moves (both doing the same dance move but without much other consideration to matching) but after a few minutes followed the approach followed by the other participants of staying relatively still and slowly moving the arms. From around nine minutes into the study they were talking quietly (as to be out of earshot) to each other. It had been explained that they should signal to end the session whenever they wished to do so but in the interview they confirmed that in the first session they kept on using the system longer than they wanted to because they felt obliged by the experimental context. Towards the end of the first session they seemed to be continuing almost as one would during an exercise routine.

The participants indicated that although they gave it a good shot, they did not really engage with the system. On a number of occasions they referred to their age as a factor within this, seeing this kind of system as well as the matching idea as something that young people might do but was less suitable for them. Matching or mirroring was interpreted only as leading and following. That said, their reported means of exploring was similar to others—*'trying to find the system's possible sounds'*, beginning with familiar movements and then pushing out to explore the system's perimeters. Their reported unfamiliarity with such types of system had an interesting effect on their expectations. R mentioned an expectation that the mapping be consistent, though referred to this as something implied by having had the range of inputs explained.

When reflecting on the session without the matching constraint, the participants seemed to talk more in terms of responding to the music rather than vice versa. R compared the music to what might be heard at a spa, framing it in terms of relaxation rather than movements and comments (*'it is not something I would have chosen to play for moving to'*). She said she was *'responding to the music'* and that she needed to be *'inspired by the beat'*.

Even after having clarified after the first interview that the participants were free to stop whenever they felt like it, there was still evidence that they had continued through obligation (R: *'I was thinking "Oh God I want to stop!"'*).

8.4.6 Participant pair 6

Participant	Gender	Musical skill	Dance skill
L	F	3	2
R	F	2.5	1

The participants did not have the mapping explained to them before they began. Due to technical difficulties (described below) the first session was restarted 3 minutes and 6 minutes in. Following this final restart the remainder of the first session lasted 19 minutes. The second session was uninterrupted and lasted 18 minutes.

Findings

Both of these participants were in their early twenties and researchers in the field of digital music signal processing. In this session there were a number of technical difficulties with the skeleton tracking, leading to the participants being asked to recalibrate a number of times and the first session being restarted twice. Prior to the second restart, R was asked to change from a skirt into trousers, which somewhat improved the tracking but it was still less reliable than in other sessions. They were also asked not to touch each other as this has been known to cause issues. Although the session was not subsequently restarted, the participants were interrupted and asked to recalibrate a number of times. Counting from this final restart, the first session lasted 19 minutes. The second session was not interrupted and lasted 18 minutes. Although the participants recognised but did not know each other previously, they established a good rapport early on and were quickly smiling and laughing.

Both participants indicated a strong preference for the first session. They reported finding it more difficult to find out what was going on in the second session as well as a greater sense of agency over the sound in the first session (*'In the second one it seemed like the sound was already playing . . . whereas the first time it felt like we were actually making the sound'*). Both participants estimated inaccurately the length of the first session but in opposite directions. L initially estimated 30 minutes and R 5 minutes. After negotiating they compromised to some extent with L saying at least 15–20 minutes and R *'maybe 7 or 10 minutes'*.

Their main comments on the matching constraint were that it assisted them in identifying the source of the sound (particularly in terms of which participant was responsible). In the second session, a lack of ability to make sense of why sounds were coming made it seem to L less systematic and more random. She contrasted this with the first session that felt more *'like it had a beat'*. They also mentioned the disinhibiting effect it had by creating a sense of solidarity. But L did also mention that in some ways she felt safer in the second as she did not have to worry about moving in a way that L could follow.

8.5 Overall Findings

We will here draw some more general findings from the above results in the context of the earlier theoretical content of the thesis.

8.5.1 Perceived agency

As mentioned in Section 8.1.1, the second session was not a control in the empirical sense. The mapping of IA was built around the matching constraint. In contrast, in the second session the mapping was calculating a mean position of their joints. As a result, the system seemed less responsive and it became difficult for participants to identify their own contribution. This was borne out through frustrations from the participants.

However, even though participants perceived more singular 1 agency ('I') without the matching constraint, their language indicated a much stronger dislike of this second type of lost control.[2] This provides tentative support for our proposition in Section 7.3.2 that first person singular agency may

[2]The 1, 2, 3 notation for perceiving agency is defined in Section 6.4.3.

be replaced with first person plural agency without appearing 'unjustified'. However, where the loss of agency is attributed to the ineptitudes of the system, it becomes more frustrating. Also of interest is that the two participants with the most musical experience (and who seemed to respond most positively overall) reframed these apparent ineptitudes more positively within the glitch aesthetic (see e.g. Gurevich and Treviño (2007) for a discussion of glitch).

8.5.2 Exploration

As stated in Section 8.4, the above analysis subsequently informed the model of emerging structures presented in Chapter 4. This is demonstrated in the approach of the interviewer who continually attempts to distinguish between participants that followed a broad (horizontal) or deep (vertical) approach when exploring (in line with Gurevich et al. (2010)). However, although participant pair 4 (Section 8.4.4) indicated that they began broadly and then moved onto *'technical'* aspects, the other participants did not seem to respond to this. This is most notable with pair 2 (Section 8.4.2) where the interviewer attempts two or three times to bring up the distinction but the participants refuse to make use of the concept and start resorting to very brief answers (*'nah just exploring'*).

An alternative description of how the system was explored was given by R in study 4: finding *anchors*—safe points to return to (perhaps comparable to Arfib et al.'s (2003) *rest points*)—and then venturing from there to find novel aspects of the system. This approach aligned more accurately with the reports of the other participants and contributed to the consideration of both the wider exploration of unexposed inputs and the deeper exploration of exposed inputs in the Emerging Structures (ES) model (Section 4.2.4).

The most common reasons for ending a session appeared to be frustration with mapping inconsistencies (especially when participants had set themselves a goal) and the sense of having exhausted the possibilities of the system, which aligns with the ES model Section 4.2. A number of participants demonstrated this desire to alternate between inputs that were familiar and inputs that were unfamiliar.

Differences were identified between serendipitous and deliberate discovery of inputs. Those who were told what the inputs were seemed to set themselves the goal of finding them all. They seemed more to position the instrument as a tool to allow them to do 'work'.

8.5.3 Rights and obligations

It is of interest to analyse the *rights* that participants inferred they were granted by the context and the *obligations* these placed on the system (see e.g. Banister et al. (1994, ch. 6) for a discussion of rights and obligations). Participants who had the mechanics of the system explained to them appeared to infer an obligation of the system to respond consistently and in the manner explained. This assumption is not unreasonable when using such a system, however those who were left to discover the input themselves seemed to be more accepting of inconsistency. Obligations placed upon the system to respond consistently went hand in hand with a right of the participant to be able to *use* the system in accord with the understanding they had developed of the mapping.

A number of the participants expressed that the system had an obligation to inspire them to move (L from pair 1, L and R from pair 5).

Overall, providing an overview of the mapping to the participants, as well as previous experience with interactive systems, influenced the rights they assumed and the obligations they imparted on the system. Frustration was the typical response if these rights were not fulfilled.

8.5.4 System-specific

Overall, the response to the system was variable. The experience seemed to depend quite considerably on the rapport established between the participants. Providing some initial task to establish rapport might have been beneficial, such as R from Study 4's suggestion of starting slowly with eye contact (see Section C.3.2).

The system was too lenient in its measurement of deviation for many participants, who did not seem to be constrained enough. It is difficult to know the extent to which this technical issue compromised the matching constraint. Whilst most participants were still trying to match each other even with this leniency, they might have been made to explore in a more focused manner otherwise.

8.6 Discussion

> Test subjects should be treated like the valued patrons of a restaurant. Listen carefully to what they say. They are experts on what they like. But let the chef adjust the recipe. Subjects are not designers. Their suggestions are valuable, but designers need to think about the results carefully. (Clanton 2000)

For games at least, it can be difficult for the users to explicitly identify what makes a good game (Sykes 2006). However, they are often quite capable of identifying things that are nots enjoyable. What an individual *wants* and what they *like* are distinct concepts (Litman 2010). In the business of exploiting anticipation, ambiguity and curiosity to create a specific experience as we are, specifically denying your audience what they want may well be the means to create an experience that they like. Thus, it has been essential to analyse our user responses beyond the surface of what was said.

In this study, our success at creating a captivating interactive music experience has been variable. Whilst developing rapport between participants seemed to play an important role in allowing them to engage with the system, it was not enough (e.g. Study 3, Section 8.4.3). Perhaps the most valuable insight that arose from this study is the role that expectation and perceived rights played, as well as differences in task-driven and random exploration. Those who felt that the system was setting them challenges seemed to run out of steam as soon as they could not perceive any new challenges. However, those that were setting their own challenges were more content to accept the system as it was—something to be *explored* rather than *utilised*.

As mentioned in Section 8.4, both the Emerging Structures model (Chapter 4) and the framework of Perceived Agency (Chapter 6) were refined following the findings of this study. Therefore, although the findings offer supporting evidence for both theoretical contributions, it would be circular to consider them a *validation*. Such an outcome is common when conducting qualitative research in this field (see e.g. Edmonds and Candy 2010) and does not indicate that the findings themselves are any less meaningful or reliable. We will return to this topic when considering future work in Section 10.1.2.

8.6.1 Reflections on the method used

Analysis of the interviews took two to three times longer than anticipated. From recordings to conceptual map took around a ratio of 90:1 analysis time to interview time. Stowell (2010) reported a ratio of 40:1 for his analysis, however this excluded the time taken to transcribe interviews. There are nonetheless a number of differences that may have made our method longer than Stowell's:

1. Our interviews involved multiple participants. Although the amount of talk would not have been significantly more, transcribing overlapping speech is time consuming. However, overall we would estimate our transcription to be less time consuming as in contrast to Stowell we did not annotate pause lengths.

2. Some of our interviews were very long and creating the conceptual maps became a lot more difficult (the complexity of clustering and constructing the graphs did not increase linearly with the number of predicates!). Furthermore, our approach to constructing these networks differed from Stowell (2010).

3. Our analysis included an additional step of resummarising the interview.

On the other hand, Stowell's analysis involved identifying different conceptual 'worlds', which our analysis did not.

More care should have been taken to ensure the interviews were kept to a reasonable length. At the same time, some participants had very little to say whatsoever. Those in the pilot studies were not especially shy and so this issue was not identified before the interview approach was decided.

Our method involves spending a considerable amount of time extracting the 'meaning' from the discourse. This approach gave the analyst a very close understanding of the text, which greatly informed further development of the theoretical contributions of this thesis.

When considering the analytical output, extracting predicates may appear to be somewhat contrary to the spirit of DA as an analysis that explicitly *avoids* dealing with the mechanics of the text. However, reflecting from the analyst's point of view, the process of extracting even descriptive predicates from interview speech required significantly more conscious analysis of the discursive mechanics than would typically be arrived at through close reading. It should be emphasised that whilst the time-consuming analysis of the discourse required subjective decisions over how predicates were formed, all findings and conclusions were drawn after the analysis was complete. A lengthier analysis time does not therefore provide greater opportunity to uncover more subtle effects.

Overall, we are inclined to conclude that the method was successful. However, the time investment involved in following this analytical method was considerable and it is not recommended that it be undertaken lightly.

Chapter 9

Discussion

In this chapter we return to some of the questions raised at the beginning of this thesis and consider how this thesis has contributed. This discussion is framed around two issues in particular: what makes for an interesting interaction and how can we describe it in a way that captures some of the richness of our everyday experiences?

9.1 What *is* interesting in interaction?

We began this thesis with a quote from Edmonds et al. (2006) asking what exactly is interesting in interactive experiences. The research presented here suggests that discovering an ability to perceive, influence and control one's environment is one factor. However, there are important considerations. In Chapter 4, I presented a model suggesting that what sustains interaction is the anticipation of further discoveries. In Chapter 6, I then argued that both when participating and interacting, the value of such experiences is mediated by the extent to which skill has been demonstrated, which is itself determined by the perception of agency—the freedom and responsibility provided by the context. The challenge for interactive artists is therefore to design systems that provide continuous scope for a participant to develop a perceivably personal ability to learn and manipulate a system.

The above could be applied to a game and any other interactive system. We, however, have focused on interactive music. I have endeavoured to demonstrate that interactive music is not simply music that happens to be interactive. Interactive music is the participatory dual of the composed instrument as it was considered in Section 3.3.1—an exploration into sound, gesture, and how these relate. Musical interaction is in a sense closely related to dance and musical audio in being an abstract medium of expression. However, the use of sensing and computing technology introduces causality. As a medium it remains non-representational but lends itself to explorations into how we relate to the world, understand it and control it.

In combining both the experience of causing effect and perceiving our environment, we have drawn upon two analogous theories: Csikszentmihalyi's (2002) theory of flow which suggests that enjoyment is derived from an increasing ability to *act*, and the expectation-oriented models of musical perception (Section 2.4.3) which suggest that enjoyment is derived from an increasing ability to *perceive*. The Emerging Structures (ES) model of Chapter 4 describes how these two enjoyments are fundamentally interrelated through the act of exploration. Both of these are examples of the pleasure of learning. However, the ES model suggests that what we expect to find out is as important to our enjoyment as what we do find out. In terms of the curve (Section 2.2.12), we need not only to ascend the curve but to see where, with just a bit more persistence, we might ascend to. Likewise with the Wundt curve (Section 2.4.3), the hedonic value of a stimulus may be maximised by balancing simplicity and complexity. But for sustained engagement, expectations must be established in the viewer that this balance will be maintained.

We have approached this and many other issues within this thesis from a subjective stance. Whilst understanding events in terms of how they are experienced makes for tricky engineering, as we saw throughout Chapter 6 many of the effects researchers within this field are trying to achieve, such as interest, pleasure and aesthetic appreciation, are inherently subjective. I suggest that too often these are translated into objective engineering goals without a thorough examination of how they arise. I hope that the material presented herein will assist others in determining what they are trying to achieve, why and how best to go about it.

In both the constraint unlocking study (Chapter 5) and the evaluation of Impossible Alone (IA) (Chapter 8) we saw that the experience provided by a system depends not only on its affordances and capabilities, but how the participant makes sense of these aspects. Considering how a system is subjectively experienced has required us to consider not only how interaction is perceived but how an individual draws upon their wider contextual knowledge to make sense of it. A participant's expectations of what they *should* be able to do is important, as are their beliefs of what others would have done in the same position. This has direct practical consequences for those creating interactive music. For example, the findings of Section 8.5 suggest that simply explaining to participants how a system functions may alter their goals when using it and their interpretation of its behaviour.

We addressed this issue from the perspective of perceived agency (Chapter 6 and developed a framework to theorise about how personal value is attributed to the phenomenological experience of spectating or interacting. The importance of considering how different types of agency is perceived and in whom was justified through identifying universals of existing musical practices and highlighting how new contexts of performance and interaction may fall short. One area where this may be of practical use beyond those creating interactive music is that of DMI creation and performance, where it may be used to identify what types of skills are expected of a performance context and whether their respective types of agency will be perceived. However, the framework is intended as a tool to question and understand the role and impact of technology within music rather than to impose hegemonic requirements on musical practice or judge what 'is' or 'is not' music. In particular, we have seen that there is not a 'correct' distribution of agency that must be perceived but rather that problems may arise when expectations of agency are inferred from a context which are not then demonstrated.

The *uncanny valley* describes an observation that as a robot increases in realism it becomes more familiar (Mori 1970). However, as it approaches a point just short of being realistic it rapidly becomes disturbingly unfamiliar. It is interesting to consider this in terms of the perceived agency of a system. We might argue that as an IMS becomes more similar to a musical instrument, although remaining clearly outside this category, it becomes more familiar and pleasing to interact with. However, there a point in this progression may emerge where the system will be similar enough for a participant to begin inferring expectations of it granting the same degree of agency as an instrument. Should the IMS fails to live up to these expectations, it becomes frustratingly unaccommodating and falls into the uncanny valley.

Whilst the ES model is a simplified abstraction of how a participant may explore a system, it allowed us to derive practical design principles: *incremental constraint unlocking* (Section 5.1) and its successor the *implicit constraint* (Section 7.2). Evaluating them is difficult as any implementation will involve a creative work which introduces its own effects. Our approach in testing the latter was to allow artistic goals to drive the development of IA—as we imagine they would in any other 'real world' application—and consider in detail how adopting the principle interacted with these goals and shaped the final result. The detailed qualitative analysis conducted in Chapter 8 then provided an understanding of how the many facets of a participant's expectations, understanding and interaction with the system combine to form their overall experience.

In Section 1.2 we chose not to restrict what types of system we were interested in, seeking instead to focus on the experience of interacting with an arbitrary IMS. We did, however, limit the scope of our enquiry to non-expert individuals interacting with an IMS for the first time. We may reasonably assume that any such experience will involve the participant learning how their actions affect the system (even if they learn that their actions do not appear to have an effect). Whilst the dynamics of an inexperienced participant forming an understanding of a system for the first time is not necessarily the concern of all IMS designers, I propose that both Emerging Structures and Perceived Agency may be applied as a means to analyse and guide the development of any such system.

This generality arises through the abstract and subjective nature of both theories. Therefore practical design principles derived from the theories may not be so universal. For example, incremental constraint unlocking attempted to guide the emergence of structures in a specific manner through encouraging participants to explore deeply before broadly. As discussed in Section 5.6, this may require a certain degree of overall complexity within the system and limitations over how and when it is communicated. It may also disrupt the perception of agency making it more suitable for game-like systems where a participant is not expecting musical decisions of the designer to become apparent midway through the interaction (see Section 6.8.2).

9.1.1 Relationship to Speech Act Theory and Musical Acts

In Section 2.4.1, Speech Act Theory (SAT) was described as a means to reason about agents seeking to alter their beliefs through effective utterances called speech acts. SAT was then applied into the musical domain as the theory of Musical Acts (Murray-Rust and Smaill 2011, see Section 2.4.1).

The ES model is similar to SAT in that the participant forms intentions based on an overall desire—either to maximize capacity or minimize uncertainty. In this case, similar to Herzig and Longin (2002), the desires of an agent are defined with respect to their beliefs. Following this analogy, the output of the participant might be considered equivalent to the utterances of speech acts in that its performance can serve both the direct purpose of creating a sound and the indirect purpose of providing information to inform the participant's beliefs.

However, Herzig and Longin's application of SAT to allow agents to adopt new beliefs relies on cooperation between agents. An agent A will communicate an intention to acquire a particular belief; B, being cooperative, will then provide an assertion regarding this belief. Similarly, in Musical Acts, an agent will seek to change the commonly agreed musical state through playing music corresponding to a different state—but the efficacy of this action relies on other cooperative agents interpreting this intention and acting accordingly. In contrast, the ES model reasons about participants behaving without a presumption of cooperation from another agent. We do not assume that an IMS is programmed to respond with the intention of altering the participant's beliefs about its future behaviour. The focus of the ES model is on how the desires of the participant affect how enjoyable their experience is, the participant's ability to form and perform intentions that will allow them to realise this desire, and how the prolonging of intentions can alter the perceived value of the belief in question once its adopted.

There is, however, a clearer relationship between SAT and the framework of Perceived Agency. As we saw in Chapter 6, a performer needs to perform not only the 'direct utterance' of a particular musical output but also the 'indirect utterances' communicating their intention to create this specific output and the parameters of agency within which the act is taking place. These three meanings may then be combined to infer the belief among an audience that the musician is skilled. I argued in Section 6.1.3 that the need for this latter indirect utterance is perhaps a peculiarity of musical acts where the audience is unfamiliar with the musical interface being used.

9.2 Putting theory into practice

As mentioned above, both the ES model and the Perceived Agency framework are most applicable in the conceptualisation, design and composition of IMSs, as well as informing the context of their use.

9.2.1 Emerging Structures

The ES model suggests that an IMS should be designed not as a static product but in terms of how an individual forms an understanding of it. Techniques such as storyboarding (see e.g. McQuaid et al. 2003) are helpful in breaking up a canonical interaction trajectory into segments that may then each be analysed through the lens of ES: How much understanding has the participant formed? How much has their experience so far suggested that there is more to understand, and that this further understanding is attainable through continued use? How far has this understanding allowed the participant to do things? Is the experience so far likely to have formed expectations of a further increase in the participant's capacity to act through the system?

This process underlay the development of Impossible Alone. With the fixed testing schedules provided by the workshop schedule (see Section 7.3.4), we were able after each development iteration to reconsider how the interaction design would lead the participant along a narrative. For example, we considered what kinds of exploratory movements matching individuals are likely to try initially—arms then torso then legs—and whether the system would reward this kind of early behaviour with new musical capabilities.

The ES model is therefore used to inform a temporal approach to design. It is also applicable in how a work is presented. We have seen suggestions both in Section 8.5.2 and from Costello (2009) that providing too much instruction can reduce a participant's enjoyment of an installation. However, applying the ES model, we see that the pleasure is as much in *becoming able* to use something as it is in its actual usage. This is in contrast to Costello and Edmonds (2009) who argue that working out how a system works is a distinct antecedent of engaging with a system. Therefore, adopting the ES model allows us to form clearer goals of what the provision of instruction should achieve. Rather than telling the participant how something works, instructions should be just enough to inform them that there are interactions to discover. Where aspects of the interaction design may prevent an individual from forming anticipation of a work's potential, as for example the matching constraint did in Impossible Alone (Section 7.3.2), then instruction may be necessary to overcome this hurdle. But for long-term engagement it is important to communicate that only a fraction of what is possible has been described.

9.2.2 Perceived Agency

The key practical applications of the Perceived Agency framework also regards communication. The framework suggests that the agency perceived to be granted by an IMS needs to match that which the participant believes is implied by the context (Section 6.8). Whilst there is a spectrum of possible configurations of first and second person agency within IMSs, failing to live up to the participant's expectations risks frustration, annoyance and dissatisfaction. This need to balance expectations with outcomes should inform communication with the participant. This includes preliminary material such as any provided instructions or description of a work—words such as 'instrument' carry implications. The framework may also inform how feedback is designed. Whilst we saw in Section 2.2.7 that feedback is important for a participant to learn how to control a system, we see here that it plays the more fundamental role of proving that a participant *is* controlling a system, and that there is the potential to greater control in the future.

The framework provides the means for designers to consider more closely *why* a work they are creating should be interactive. Jordà (2004, p. 324) argued that 'faked or useless' interactivity degrades contemporary art. The Perceived Agency framework helps avoid this fate by demonstrating three categories in which interactivity allows a participant to take ownership of an outcome. There will be other factors shaping the perception of agency; however, those creating interactive work requiring no scope to demonstrate skill or originality or for expressive communication may need to consider more closely what exactly they are hoping to achieve with interaction. Does the provision of interactivity allow a participant to make an experience unique to themselves? This is not a question that necessarily needs have a positive answer but it is an important aspect of how interaction affects an interaction in need of consideration.

To demonstrate how this may be applied in practice, we might consider a work that produces sound through a brain interface (e.g. Electroencephalography (EEG)) provided to a participant. Applying the Perceived Agency framework we see that if the interface is presented as an interactive installation but does not create the perception of first person agency then it may create a dissatisfying experience. For such an interface, expressive, technical or creative first person agency may be technically unfeasible. However, referring to the framework we see that we may still establish ('loose') first person agency by providing a sense of uniqueness. This suggests that if we were able to design the system so that it responded to each participant in a unique but consistent manner, we may sidestep the challenge of providing perceptible control. Note now that seeing the interface in the hands of others forms a fundamental part of the participant's experience, introducing additional requirements upon the context in which the system is presented. Having defined more clearly the relationship between participant and system, we may observe that describing the system as allowing the control of sound through thought may create false expectations. Perhaps better would be a description of the system as providing a window to experience the existing signals eminating from the brain.

9.2.3 Personal reflections

The work within this thesis has led me to identify and reappraise many of my assumptions about why I—and others—are drawn to creating musical interfaces. I began from an engineering perspective of seeking to solve a problem but found it increasingly difficult to disentangle motives surrounding usability and expressive potential with those of my own musical ambitions. Conversely, I presumed that creating music and participating in musical acts were purposeful ends in and of themselves. Of course, music and art need no utilitarian justification. However, to examine why some musical interfaces failed to satisfy the musical 'urge', I found it necessary to consider more closely the origins of the desire for musical expression. Although I regarded my own musical practice as often being quite solitary in nature—both in terms of creating and playing instruments—social action and communication appeared to underlie nearly all musical practices much more that I had considered. Many are understandably cautious about reducing aesthetic experiences to scientific 'causes' or less appealing aspects of human behaviour such as status seeking. However, I found that rather than explaining away the beauty of music, the more I examined its origins the greater the presumption that music is primarily about sound appeared something of a hegemonic restraint on interactive music (Small (1998) argues this point in more detail).

The consequences of this shift in attitude may be seen with Impossible Alone. By considering music as a social experience rather than purely an exercise in sound creation, I was able to concern myself less on ensuring a participants would be pleased with what they heard and instead focus on how the installation and the context it established would allow two individuals to form and explore their relationship with each other.

Whilst these are reflections on a personal journey, it is possible to draw more general lessons. As we saw in Section 2.2, the field of instrument design often measures its success through comparison with traditional instruments. Whilst traditional instruments were designed for a range of considerations, the most obvious is surely to produce a specific sound. With digital instruments this is no longer the case. Even if a new musical interface is created for performance in the concert setting traditional to Western art music, in the vast majority of cases, the sounds could have been composed and pre-rendered beforehand (see Section 6.1.4). However, digital musical instruments allow for new musical events—the realtime control of a digital sound—and consequently opens up possibilities for new types of events such as this sound playing a role in an interactional situation. Whilst a DMI's purpose remains to produce sound, the need for it to be designed is often to change the context and social significance of the performance of certain types of music. This suggests it is perhaps a mistake to think of new musical instruments as entities divorced from the context in which they will be used. A musical instrument is not just a means of providing control over a sound space. It allows an individual to *perform* and so demonstrate ability. We saw in Section 6.5 that different musical practices emphasise the importance of different types of ability. If an instrument is to succeed in a particular context, then its design will need to allow this ability to be made evident.

9.3 Reflections on the research process

A PhD is a journey in which both the research question and the process through which it is investigated are evolving throughout. With an interdisciplinary topic such as ours, it is necessary to consider at each stage not only what to investigate but also which research methods will be most suitable in doing so. It may therefore be helpful to future researchers to briefly reflect on some of the decisions that were taken and what I might have been done differently with the benefit of hindsight.

The research question of this thesis, *how can interactive music be made more captivating?*, is broad and open ended. I decided in Section 1.3 to allow the specifics of what interactive music is, and how we might define and measure captivation, to arise naturally over the course of the thesis. A strength of this approach is that it allowed us to draw on a wide range of disciplines, including psychology, musical perception, Human-Computer Interaction, New Interfaces for Musical Expression and interactive art. We were able to identify common themes in how these fields address engagement and consider how these might be applied more specifically to interactive music. However, this broad approach raised difficulties. Without declaring at the outset a specific indicator of success, following an engineering methodology became more difficult as we were faced with the concurrent tasks of identifying our problem, solving it and developing a convincing objective indicator of our success. In retrospect, this was perhaps too ambitious. This is a long thesis yet there has not been the opportunity to validate its theoretical contribution empirically to complete satisfaction.

We saw in Section 1.3 that the goals of interactive art are often subjective and so addressing subjective requirements is a necessity when considering the engineering challenges of interactive music. I maintain the position that subjective models and frameworks such as those of Chapter 4 and Chapter 6 provide an effective tool to do so. However, mixing subjectivity and engineering requires a diverse canon of research methods and greater consideration should have been given as to how quantitative and qualitative methods may be used complementarily rather than as alternatives. It is possible that had the qualitative study of Chapter 8 been conducted earlier, a better targeted quantitative study may have led to hypotheses that faired better than those of Chapter 5.

Ultimately, however, I am glad to have had the opportunity to draw upon such diverse fields of research. And whilst the findings of Chapter 5 were not statistically significant, understanding why this was the case formed the basis of the subsequently more successful evaluation of Chapter 8.

9.4 Types of interaction

In Section 2.1 we reviewed a number of attempts to describe different types of interaction such as discussions of 'conversational' interaction (Section 2.1.3). We shall consider here how this distinction might be better thought of in terms of perceived agency.

Many of the notions we met in Section 2.1.3—mutually influential (Jordà 2005), a two-way exchange of ideas unique to a pair of participants (Paine 2002), cognition and memory (Bongers 2000)—are attempts at objectively defining what makes conversational interaction special. However, we have been arguing that interaction is more usefully considered subjectively from the perspective of the participant. From our day-to-day experience, we are able to determine the difference between a conversation and simple non-conversational exchanges of words based entirely on what we perceive. We do not need privileged insight into the subsequent state of mind or levels of cognition of the other party in order to determine whether we were conversing or they were simply providing scripted replies. I will here propose the position that for an interaction to be perceived as conversational it is necessary to have perceived both 1 and 2 agency where, following the notation we defined in Section 6.4.3, 1 corresponds to the participant whose perspective is of interest and 2 to the participant's interlocutor. We shall consider these two roles in terms of 'I' and 'You'.

Recall the earlier requirements of agency: intention, forethought, self-reactiveness, self-reflectiveness. Now consider communication between two parties where

1. You appear to be, or I am, speaking without the intention of being understood by the other party;

2. You appear to be, or I am, unwilling to use your understanding of the other party in order to turn this intention into concrete plans of what to say;

3. You appear to be, or I am, unable to determine when we have or have not been understood by the other party;

4. You appear to be, or I am, resigned to the fact that the other party cannot understand what is being said.

Furthermore, including the requirement of perceived performative agency as having scope to exercise agency (Section 6.3.3):

5. You appear to have, or I have, no significant choice over what we are saying.

I argue that any of these subjective experiences are enough to convince an individual that their interaction is less than conversational. As a result, the perception of live 1 and 2 agency is a necessary component of conversational interaction.

We saw that a performance system could demonstrate live agency in Section 6.6.3 with Rowe's artificial performer. However, we also saw that Rowe's performer receives *musical* input (Section 2.2.4). Furthermore, Rowe's third classification dimension considered systems to be either in the instrument paradigm *or* the artificial player paradigm. From this we may argue that Rowe sees his systems as offering a user either the means to exercise 1 agency (the instrument) or a demonstration of 2 agency (the artificial performer) respectively—but not necessarily both. Conversation can be a misleading metaphor for describing interaction with an Interactive Music System (IMS) as it is easy to neglect the importance of first person agency. For a participant—let us call her Alice—to converse with an IMS in the language of music, the system must provide both her and her interlocutor with the means to express themselves and therefore simultaneously provide the perception of 1 ('I') and 2 ('you') agency. (Note, we are considering Alice's point of view here so 1 agency is her perception of her own agency and 2 is her perception of agency from her interlocutor.)

We might imagine that the simplest way in which Alice might perceive this arrangement would be to construct a mental model that divides the system into an instrument (her voice) and a *modifying agent* 'hidden inside the system'. Such a construct is exactly what we saw in Section 2.1.2 with Cornock and Edmonds's (1973) 'modifying agent' (Figure 9.1).

We objected to Cornock and Edmonds's system because it was defined as an objective reality but the modifying agent was effectively imperceptible to the participant and so irrelevant in determining their experience. It is important to note that we are here considering a mental model that has been constructed by Alice in response to the system's behaviour, not an objective definition of the system.

However, whilst we may imagine a few unusual exceptions, conversations typically happen through a common medium such as a shared acoustic en-

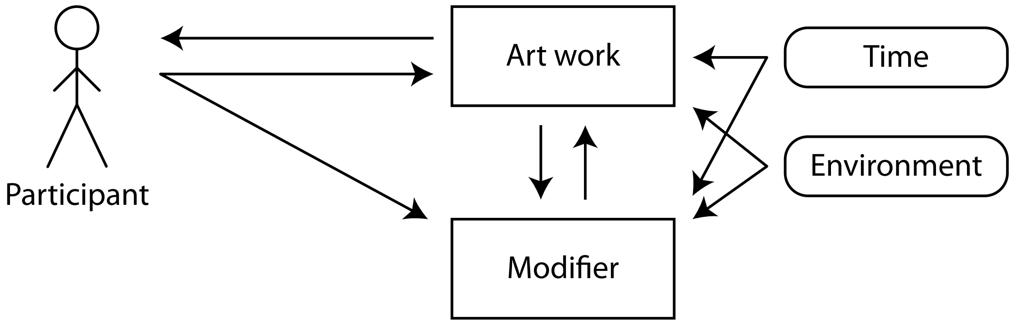


Figure 9.1: Cornock and Edmonds's dynamic-interactive system (varying). See also Figure 2.2. Adapted from Cornock and Edmonds (1973).

vironment. Such a medium is accompanied by a common understanding of how an agent may affect that medium and how this will be understood by the other party. If Alice is conversing she can hear her own voice and believes the other party may do the same (aiding her perception of self-reactiveness and self-reflectivity). Furthermore, by the need to perceive 2 ('you') agency argued above, Alice also needs to believe that these conditions hold for the other party. This produces a perceived system in which both parties are able to influence predictable control upon a common medium (Figure 9.2).

One thing we may observe from this diagram is that the output of the system mixes together both Alice and her interlocutor's influence. It is therefore essential that she is able to predict the effect of her own influence in order to disentangle these two inputs. When we converse with speech we take turns; one effect of this is that it makes it easier to understand what is said by the other party. However, note that this still requires both parties to be able to be reliably quiet when it is the other's turn to speak, something that is not necessarily possible with an Interactive Music System (IMS) without some knowledge of the mapping. In Section 2.1.2 we argued that this was silently assumed of the varying system in Cornock and Edmonds's classification.

With this perceived difference between an agent within this system and a mutually accessible medium, we are able to understand why Jordà (2005)

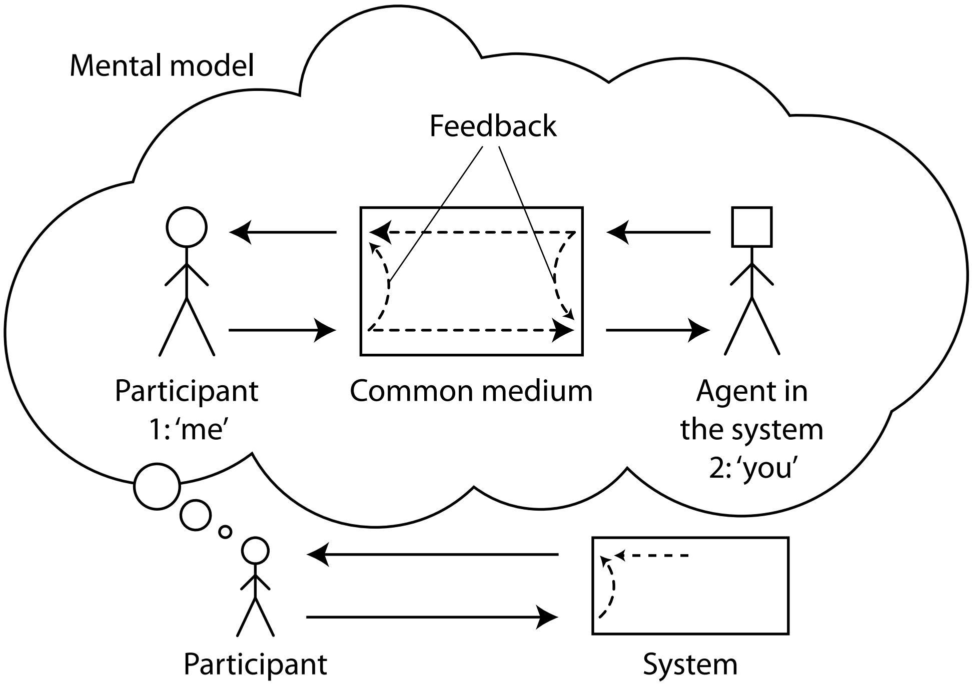


Figure 9.2: A mental model of conversational interaction.

did not think why an adaptive signal processing unit did not involve human to machine communication and why Paine (2002) thought a racing game would only be interactive if the racetrack started changing (Section 2.1.3). Whilst these systems may provide the means to exercise 1 ('I') agency, their behaviour does not provide evidence of an agent within the system.

Note that we are not saying that the perception of 1 and 2 agency is necessarily *sufficient* for conversational interaction. However, I would argue that justifications provided for cognition, mutual influence, memory, complexity or being left in a different state (Jordà 2005; Bongers 2000; Paine 2002; Dannenberg and Bates 1995, see also Section 2.1.3) are all subsumed under this requirement.

Recall Chadabe's (2005) three metaphors of interaction: the powerful gesture expander, sailing through stormy seas and the conversation (see Section 2.1.3). Considering these in terms of a mental model as a split between instrument and agent, we may more rigorously define the different experiences they describe.

The gesture expander may be seen as establishing a feedback loop that offers reliable control, i.e. one in which the participant may control the output with intention (Figure 9.3). There is no perceivable agency from the system, however it provides the participant with perceived 1 agency.

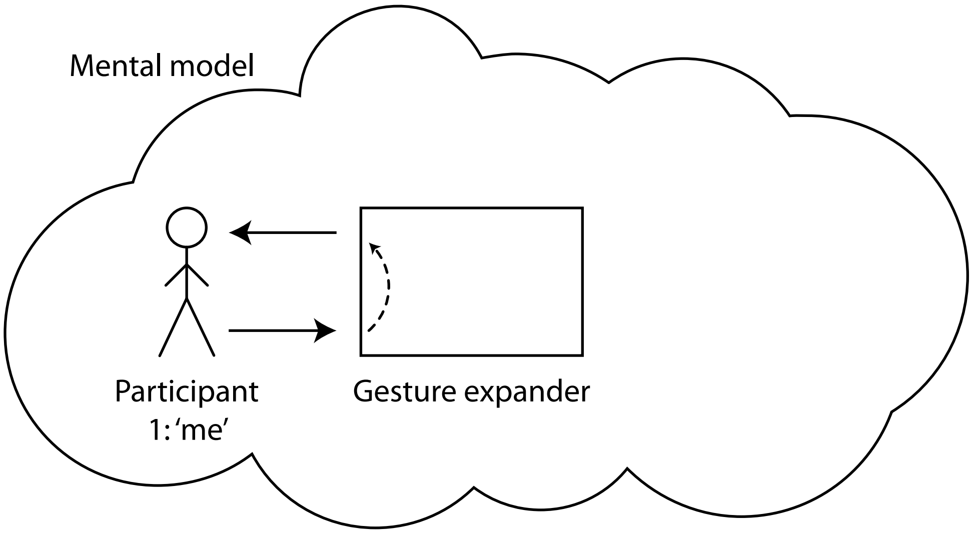


Figure 9.3: A mental model of the powerful gesture expander.

Sailing through stormy seas may be perceived as an unreliable controller where the above feedback loop is distorted in a manner that is unpredictable from the participant's perspective (Figure 9.4). Note that in order to be able to determine how much influence is arriving from the unpredictable source, the participant requires a model of how their input would be transformed without its influence (which may be estimated if the influence disrupts the input in a consistent manner, such as adding Gaussian noise). In terms of the sailing metaphor, it would be difficult to estimate the extent of a storm based only on the efficacy of the ship's controls without a prior understanding of how effective they are usually.

Chadabe's third metaphor of conversational interaction has both the participant and an agent within the system perceived to be influencing and

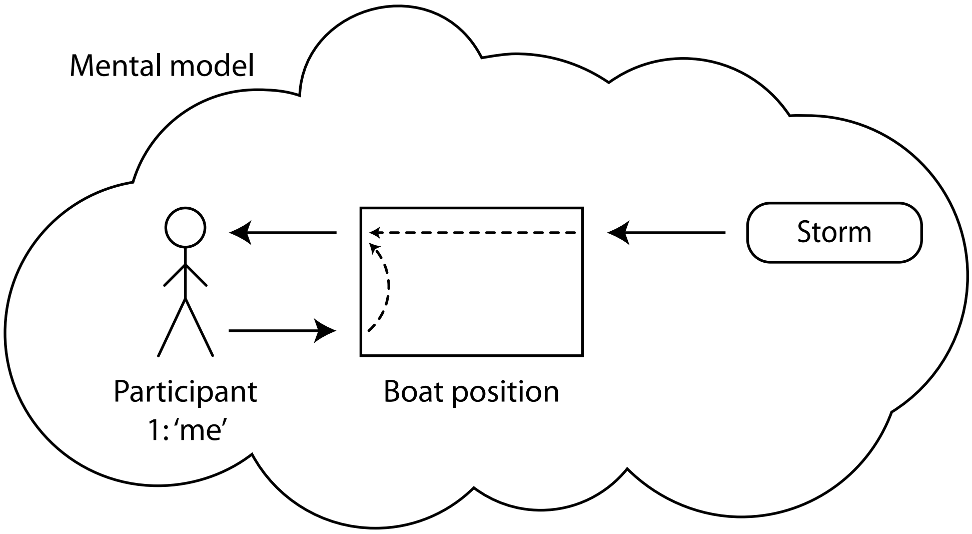


Figure 9.4: A mental model of sailing through stormy seas.

reacting to the common medium (Figure 9.5) as discussed above.

The other categorisation of interaction we saw in Section 2.1.3 was Johnston et al.'s (2009) instrumental, ornamental and conversational distinction. Instrumental involves input from the participant, ornamental from the agent and conversational from both—in Johnston et al.'s example this was seen as turn-taking between the above.

Johnston et al. present an example IMS in which participants interact with a model of a dynamic system of particles joined by springs. They argue that this system facilitates all three types of interaction: the participant could directly move the particles but when they let go they would move and then stabilise in a manner which is difficult to predict.

In order to consider this conversational in the manner presented above, we may perhaps assume that the participant is able to predict the effect they have on the particles, but not the subsequent effect of the dynamic model in which case the spring action would be the interlocutor. This would provide for a mental model splitting the participant's control from that of another party (Figure 9.6). However, the requirements of perceiving agency in the springs are unlikely to be met. Whilst we may argue that the dynamic model

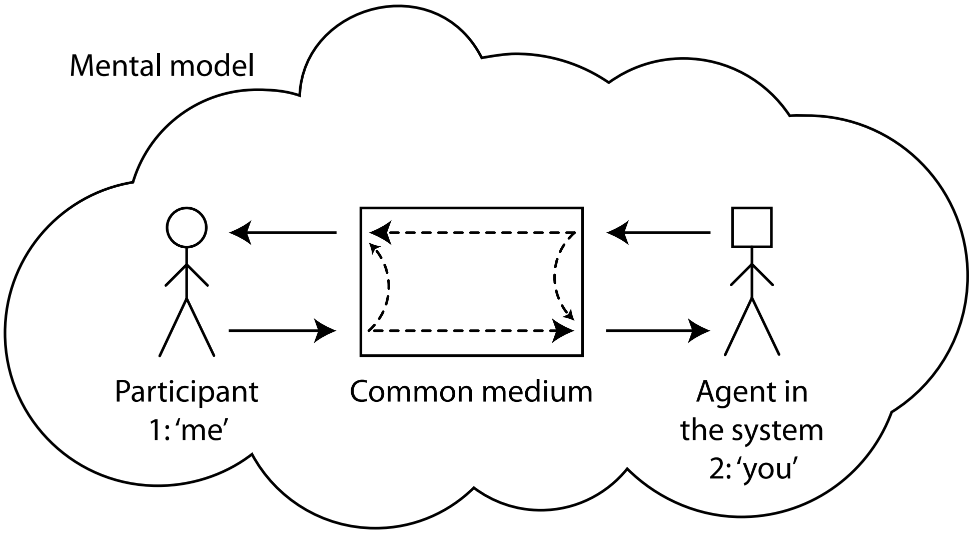
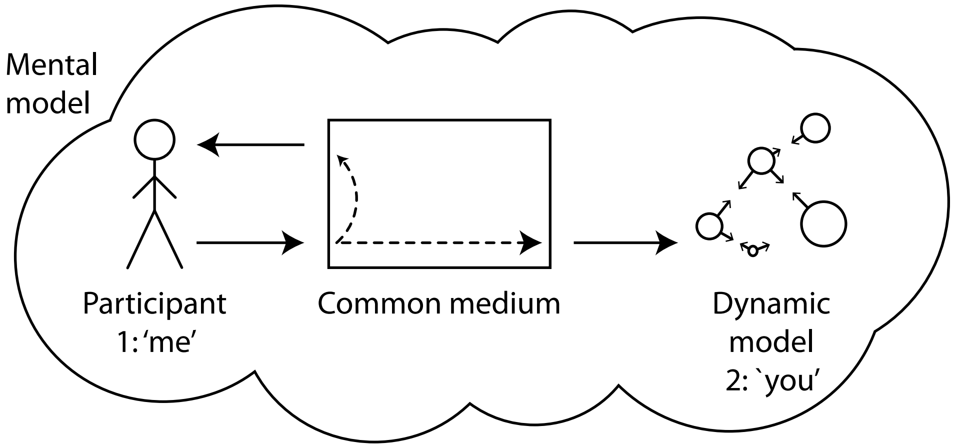


Figure 9.5: A mental model of conversational interaction.

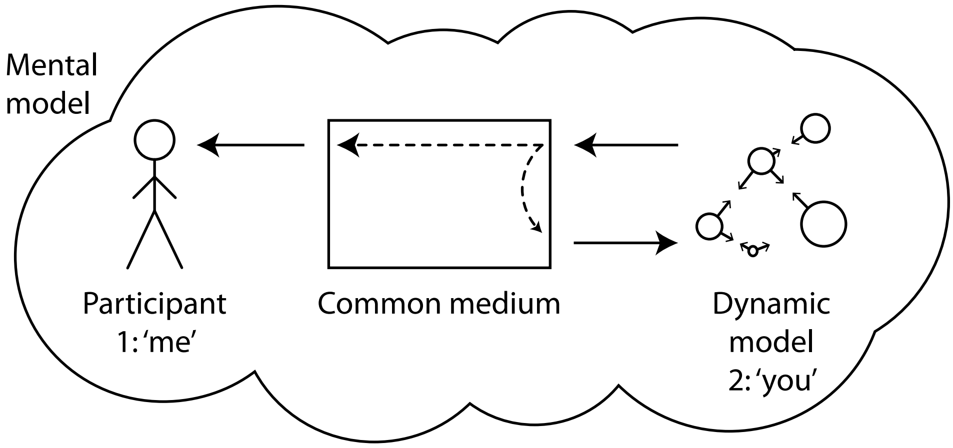
is self-reactive as the effect it has on the particles is dependent upon their current state, its actions are unlikely to be perceived as acting with intention and forethought—i.e. forming goals and executing plans in order to achieve them—or as having been chosen from a set of possibilities. As a result, we would argue that Johnston et al.'s model does not fully capture the potential for conversational systems. But it is worth remembering that the perception of agency is subjective. It may be that a participant feels that the springs are acting with intention, that they planning and choosing to apply force to the particles in order to effect a particular state upon the common medium. However, the experiences described by Johnston et al. (2009) do not suggest that this was the case.

What makes for *good* conversational interaction? For that we defer to the Emerging Structures (ES) model of Chapter 4. We may see the capacity for action, which we would like to be continually increasing, as the participant's ability to act on the common medium. We might imagine a truly captivating conversational interaction where it is through learning from and communicating with the agent within the system that this increase happens.

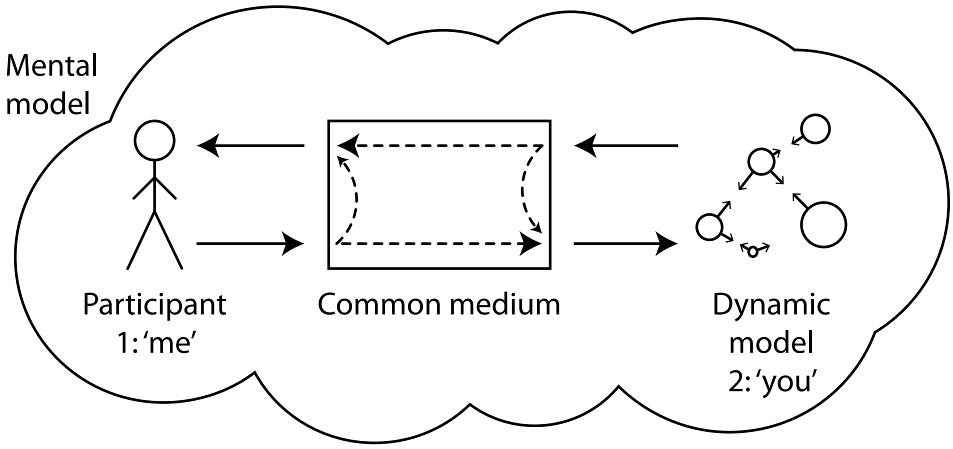
There are further paths that this discussion may be taken down. For



(a) Instrumental interaction



(b) Ornamental interaction



(c) Conversational interaction

Figure 9.6: Mental models of Johnston et al.'s (2009) categories of interaction.

example, we might argue that Chadabe's metaphor of navigating through a storm was intended less to describe unreliable controls and more to describe a system that creates a perception of agency shaping the overall musical output rather than through conversing with the participant. This might be addressed in terms of the parameters of agency (see Section 6.3.1)—an interlocutor exercises agency over communicating with us whereas such a storm would be exercising agency over the common medium. My intention here, however, has been to illustrate how the theoretical material within this thesis may inform the understanding of interactive music systems and highlight one area in particular that it might contribute towards.

Chapter 10

Conclusion

In this thesis I have argued that captivating interactive music should provide an experience with emerging structures of musical interaction without hindering the participant's perception of agency. It is an exploration into an ever unfolding world of musical interaction, where emerging structures are how it unfolds and perceived agency is what makes us explorers instead of sightseers.

In Chapter 3 I described how musical interaction may be composed, performed and experienced and as a result argued that temporal organisation needs consideration here much as it does in conventional composition. From this I refined the concept of the composed instrument and of interactive music as temporal experiences, much like music, with interactive music being the participatory dual of spectating a composed instrument performance. It is this crafted exposition and development of sound, gesture and their relationship that has the potential to turn interactive sound into interactive *music*. It is a vision is similar to Minsky's (1981) analogy of a guided tour through the building, however instead of being shuttled through a building on a train, we are free to wander on our own terms. The task of the interactive music composer is to construct a world that makes such wanderings as musical as the train ride, without removing the sense of freedom. To assist us in achieving this goal, I presented in Chapter 4 a new model of exploration, the Emerging Structures (ES) model, that furthers our knowledge of how this world results in a subjective experience and provokes further action from the participant. In Chapter 6, I then presented a new framework of Perceived Agency to understand better what a *sense* of freedom really means, the different ways in which it may manifest and why it is important. To put these theoretical tools to work, I have presented, implemented and evaluated two new design principles: *incremental constraint unlocking* (Chapter 5) and the *implicit constraint* (Chapters 7 and 8).

Throughout, I used these models to argue that it is the possibility to learn more and do more that makes for a captivating interactive experience. Continually emerging structures are a fundamental aspect of interactive music—as they are music. But interactive music has the potential to be so much more than simply music that demands participation. It is the music *of* musical action. Through exploring the performance with the creation of the Serendiptichord in Chapter 3 we saw that structures emerge not only in what we hear but how we create what we hear. However, it is not enough to provide a system that steadily introduces complexity. From the ES model of Chapter 4 we concluded that consideration is necessary of how the participant will perceive and form an understanding of how it functions, what it allows them to do and, crucially, what it is *about* to allow them to do.

In the ES model there are two ways in which structures emerge: through exposing novel inputs and through refining an understanding of exposed inputs. These correspond to Costello's (2009) investigative and diversive types of exploration or, in the words of one participant of the Impossible Alone evaluation, *'you just go mental and start pushing buttons'* and *'go into the technical aspect'* (Section C.3.2). But they also align in a somewhat different way with Litman's (2005) two types of curiosity: an interest in finding new knowledge and a sense of deprivation in our existing knowledge, in which case individual differences may play a considerable role. This model provides a new way of describing the exploratory process which allows us to unite and apply different psychological theories to further our understanding of interactive music.

We observed the importance of considering how agency is perceived in Chapter 5 when we evaluated our first design principle based on the ES model, incremental constraint unlocking. The framework of Perceived Agency I introduced in Chapter 6 provided a new way to understand what these roles are and analyse how they interact with contextual norms.

Asking the audience to take an active role in the experience means giving them an active role to play. Agency is not just about freedom but about responsibility. If you do not have the agency to create bad music then you cannot have the agency to create good music either: it is simply how it would always have been. The trick is not to make things easy for the participant but easy enough. But as we saw in Chapter 6, agency is a *perceived* quality of an interaction. If it is not perceived, or if an unseen agent is perceived to be making the decisions, then the participant is unlikely to feel the experience is really their own. Without agency, we are simply tourists on the beaten track. However beautiful the sights are, they never quite match the hidden gems you think nobody else has found.

Conversational interaction has remained an elusive goal for many researching Interactive Music Systems (IMSs) or interactive art. With a closer consideration of the subjective experience of conversing through the framework of Chapter 6, I argued in Chapter 9 that fundamental to conversation is the perception of agency in both the participant and their interlocutor. This may be a particularly fruitful way to apply the findings of the ES model in future IMSs. An interesting person may say interesting things. But the best conversationalist makes *you* say interesting things. Likewise, the best interactive music systems will not be the ones that produce beautiful music but the ones that make *you* produce beautiful music.

10.1 Future work

A number of avenues for further work now await. I describe here a few of particular interest.

10.1.1 Emerging structures through conversational interaction

The approach taken towards creating systems with emerging structures have been focused on constraining interaction. However, in Section 9.4 we saw how the framework of perceived agency may together inform the design of more conversational interaction offering an alternative means to apply the ES model. This may be a promising direction of future research into interactive music.

10.1.2 Validating the Emerging Structures model

The extent to which the ES model has been formally validated by this thesis is limited. There were a number of directions that could have been taken following this outcome; I decided that we needed a better understanding of the social and subjective issues that determine how expectations of a system influence the experience of interacting with it. Whilst the evaluation of Impossible Alone in Chapter 8 provided supportive evidence, a number of predictions within Chapter 4 await further evaluation.

10.1.3 Further development of the Emerging Structures model

The ES model is oriented around a cognitive understanding of interaction. Although it may be applied both to conscious and embodied thought, we have not differentiated between these. Further consideration could also be given to the role of erroneous inference on the part of the participant or how their knowledge changes over time (e.g. forgetting, distortion).

The ES model at present does not consider prior knowledge in much detail or the effects of repeated exposure and vicarious learning through observing other participants. In particular, these three factors are likely to inform the participants expectations of perceived agency. For example, one participant in the evaluation of Impossible Alone suggested that she would be better able to make sense of their experience if she were able to see others use the system (Section C.5.2).

10.1.4 A computational model of emerging structures

As discussed in Section 4.9.1, whilst the ES model is defined subjectively from the perspective of a participant, its basis on the information dynamics models of Section 2.4.3 leaves open the possibility of modelling a participant using machine learning tools.

A particularly interesting possibility opened by this would be to apply a suitable machine learning algorithm to the interaction logs of participants using an IMS. This would allow an estimation of uncertainty, encounter entropy, implications (expected decreases in uncertainty), capacity for action (channel capacity). These could then be compared against subjective measurements such as questionnaire responses allowing for a more formal validation or advancement of the ES model.

10.1.5 More interactive music

Finally, we have presented our notions of what interactive music could be, as well as a number of prototypical implementations to evaluate these ideas. But the true test will be whether any others creating Interactive Music Systems (or other Digital Music Systems) are able to apply these ideas to their own work, as a means both to understand why audiences respond in the way that they do or to guide the creation of new work.

10.2 Final remark

A skilled musician may arrive at a static Interactive Music System and develop their own structures. But a system for non-experts must do more. As much as we are granting the participant freedom over what music they create and how they go about doing so, the experience, their ability to see potential, their motivation to explore and learn to use the system—these all remain our responsibility.

Just as in any medium, some works will be more challenging than others and may be created for more seasoned audiences. The composer of interactive music of course remains free to create challenging work. What we have sought to establish here is the basis of the medium. Rules are made to be broken. Hopefully we have clarified some of the consequences.

Appendices

Appendix A

Further details of the percussive mapping model

This appendix provides further details of the percussive mapping model used with in the Serendiptichord, outlined in Section 3.2.4.

The percussive mapping model draws on the embodied metaphor of striking an object, which has been found to be quickly learnt by performers (Bevilacqua et al. 2011). Note however, that our model associates objects with orientations rather than positions. In this way, our approach seeks to provide with a consistent sequence of sounds to a given bodily gesture in a way that is intuitive to both dancer and audience.

Sensor inputs from both noisemakers are concatenated to produce a position within the combined orientation space. Within this space each sound object is assigned a random position when it is loaded. The object is triggered when the combined input comes within a (Euclidean) distance d of its associated orientation, with the trigger volume determined by the speed of movement towards the sound object. In this way, sound objects may be seen as spheres of radius d within the controller space that the noisemakers 'hit'.

When the noisemakers move away from the sound object, the sample is stopped when a boundary slightly greater than d is crossed to provide a slight sense of 'stickiness' to the objects and prevent rapid triggering around the

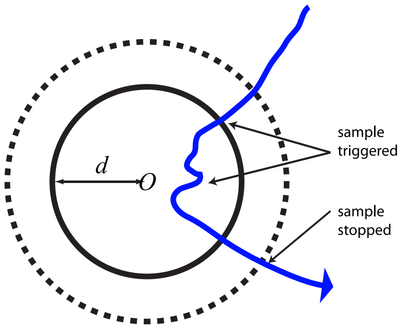


Figure A.1: A dancer's trajectory (blue line) through the controller space and within the boundary of a sound object with centre O. The sample is triggered once from entering the boundary and again from returning back towards the centre having started moving away. The dotted line shows the stopping boundary.

boundary. In initial testing, however, dancers found it difficult to predictably trigger a sample twice in a row as they had no way of explicitly knowing where this boundary was and needed to escape it and return to retrigger the sound. This was addressed by considering the speed of approach towards or away from the sound object (Figure A.1). Thresholds were set requiring a minimum speed of approach to trigger a sample and allowing retriggering once a minimum speed of departure had been reached.

Appendix B

Further details of the constraint unlocking experiment

B.1 Assessing musical experience

Participants' musical experience was assessed prior to selection by asking applicants to answer the following questions in their responses.

1. Do you play an instrument? Yes / No
2. How would you rate your proficiency? Beginner / Intermediate / Semi-professional / Professional
3. Have you ever composed your own piece of music? Yes / No
4. Have you performed music to an audience within the past five years? Yes / No

Participants were given a point for each 'Yes', and a score in the range 0–3 for their answer to question 2, giving a total score in the range 0–7. Within two days of the recruitment email being sent, 57 responses had been received and only these applicants were considered for the study. Their mean score was calculated as 2.7 and so those with a score ≥ 3 were allocated to the musically experienced group (25 applicants) and those with a score

≤ 2 to the musically inexperienced group (29 participants). Three applicants did not answer the questions and were not further considered for participation.

B.2 Statements used in the questionnaire

The following table shows each statement of the questionnaire alongside the hypothesis that it was measuring. Each statement measured either a positive point for the hypothesis or a negative point. Negative points are shown with minus signs.

Q	The best output was produced
T	I connected with the system
–Q	I did not feel satisfied with the result
A	I expressed myself the most
D	The system had a broad range of musical possibilities
–L	I was concentrating hard but did not make any progress
L	I understood what was going on
E	I was completely involved in what I was doing
D	Other people using this would probably sound quite different from me
Q	I would like to listen to a recording of the music
–A	The music was determined less by me and more by the system's creator
–T	The connection between the music and my actions was difficult to understand
–E	I was not entirely focused on what I was doing
–F	I didn't enjoy myself
–A	It wasn't my fault when it didn't sound good
–L	I learnt how to use the system quickly
–T	The interface was intuitive
–Q	I did not like how the music sounded
–D	The system has a limited range of musical output
–E	I put in effort to achieve particular results
–D	The sound was overly repetitive
–F	The experience was not worth the effort I put in
F	I would have liked to have used the system for longer
A	I was creative during the session
–L	My ability to use the system improved during the session
–F	I would like to use the system again
–E	The experience was not engaging
–T	There was not a close relationship between my input and the systems output

Appendix C

Further details of the evaluation of Impossible Alone

C.1 Participant pair 1

The mapping was explained to the participants.

C.1.1 Observations

Overall, the participants did not seem to establish a good rapport with each other.

First session

The participants did not talk except for a few quiet words three minutes in when they were trying to reproduce a specific sound. They made eye contact fairly early on and were making quite large movements early on. R seemed to be leading for most of the time although L led at one point as well. They asked to stop the session after 5 minutes, a move which seemed to be initiated by R.

Second session

The beginning of this session was delayed by a few minutes due to calibration difficulties with the system. In this session, the participants' movements seemed somewhat more repetitive. To some extent, L still seemed to be following R as before. At seven minutes into this session, both participants became quite still and focused on small slow movements. During this, they seemed to be exploring more deeply than before. The session was stopped after 11 minutes, led by R.

C.1.2 Reconstructed summaries

First interview

Participant L I found it interesting but found it difficult to catch the music. Sometimes it felt the sound was always the same but I haven't too much knowledge about music so maybe it's a little bit hard for me to know how to do actions that match this music. Sometimes after the music has finished I may think of something. I think we learnt quite quickly how to use the system. We can see each other so sometimes we can feel all 'what should we produce' and then I can produce the same actions. Perhaps with our previous experience we can come to the same idea for a similar sound. Maybe with a different partner we can try to have different actions or something.

Participant R First of all I was trying to get some kind of communication going with you, trying to work the system together. In the beginning I wanted to find out the broad range of what the system could do. Then I was trying to see which kind of sounds we could produce with which kinds of movements. Then after, I think we tried to recreate specific sounds and get more into how the thing works and see what we could do with it.

It felt like we were trying to take the same kind of approach. I was trying to work out if I could reproduce certain sounds or make a stronger connection between certain actions and certain sounds. Sometimes we produced new interesting sounds but we couldn't work out what we did to produce it.

The first thing that became clear was what the not-in-sync sound. That was fun and worked really well. That was the easiest thing to learn: when we were in sync and what sound that produced, as opposed to specific movements that would produce a specific sound. The nicest bits were when R did something new and then I was like

'Ah yeah let's do that'. These collaborative bits where eI forgot about the study were most enjoyable. It could be interesting to try it with somebody else.

I would have liked there to be more differentiation between the sounds. I found it quite difficult to know which actions would produce which sounds because there wasn't much differentiation between the sounds. It would have been easier to learn if they were more distinct. I was a bit frustrated because I would have liked it to be clearer. I learnt vaguely which actions produced which sounds but not enough to my satisfaction.

I was trying to remember what kind of movements you said you could do. I wasn't sure if I had exhausted all of the possible sounds I could produce so I'd be interested in trying to find out what other sounds the system could make. If you hadn't have told me what the possible sounds were, I would probably have tried a bigger range of movements because I remember you mostly spoke about the things you could do with your hands. Maybe if you hadn't said that I would have tried to use my whole body much more.

I stopped when I felt I'd exhausted the possibility of making a stronger connection between certain actions and certain sounds. And I had maybe got a bit frustrated because I couldn't do that.

Second interview

Participant L This second one was more difficult than the first one. In the first, you can conference with your partners and do the movement that they are doing. In this one, you have to find some special movements that are different from your partner.

Sometimes we found the connection between us and the music but mostly we did not know how to find the sounds. We did not know what we should do for this kind of music. I actually found the music was not for me. It has too much silence and sometimes you do not know what to do with it which I found a bit frustrating. There is music that I find makes you exciting and makes you want to move but this cannot let you feel exciting. Sometimes you do not want to move any more and it made me feel like I want to stop, I need to stop.

The system is quite good but I do not know what it could be useful for.

Participant R I found this second system both frustrating and interesting. I got more involved this time, but this may have been because we had already had the experience from earlier. But this time, especially towards the end, it was more like free collaboration. We were still working together but not relying on each other and doing our own thing - something we were perhaps afraid of. In the first one, we were kind of tied to each other. We had to really communicate and cooperate to get anything done. But the second version I felt we got more the hang of it and I understood more. We went deeper as well. It felt like there was more to discover and we created whole soundscapes that we were not able to get to in the first version. In the first session we seemed more to be trying to crack the system, energetically attempting to discover all of the sounds.

We might have got more of the hang of it because at one point we stopped moving. I do not really remember why I did this but possibly it was because I was confused and frustrated by not being able to tell which sounds I was influencing and which sounds she was influencing. So perhaps I just wanted to watch L. But I did not know that we could produce a new sound through stopping. This was something we almost discovered without trying and was just happened upon. It made us realise there was more to discover.

Matching each other had limited the musicality of the system. That said, although I thought beforehand that it would be much easier without us being together, at some point I realised that it seemed quite fruitful that we were together. I think losing the matching requirement was more the reason that the two experiences differed rather the fact we'd already had experience with it.

The system has lots of potential but I would have liked more control to finely tune the sounds. The sounds were also too similar to each other and I would have liked more variety and delineation between them. I do not think either session would necessarily have been that different if I was working with someone I knew.

C.2 Participant pair 2

The mapping was explained to the participants.

C.2.1 Observations

First session

The participants quickly seemed to establish a good communication and eye contact. They had a good rapport and were laughing although they appeared slightly squashed by the marked perimeter. L interrupted the study 5 minutes in to ask a question about occlusion. After hearing the answer, L led the participants in asking to stop the session. Throughout this session, the participants did not particularly come across the mapping of the hand speeds to the string level [something that has proven quite popular in other scenarios].

Second session

There was not much eye contact or communication. L began exploring the string volume slightly more around 2 minutes in. The participants ended the session after 4 minutes, led by L.

C.2.2 Reconstructed summary

First interview

Participant L The system is really interesting, good, nice, fun and is best described by the term 'interactive music system' although it could be a game. It could be seen as movement-based production, especially if you could load your own sounds onto it but more as a fun way of coming up with ideas rather than for something like mixing. However, it is better described as a game.

At the beginning we were mainly just messing around but then we started finding the locations of particular sounds. The sound design was really nice. I occasionally play music with other people but this was more like a game.

The matching requirement brought another dimension to the experience and this made it more interesting and exciting and meant that neither of us was alone when making use of the system. Cooperation just happened and was funny but good with the matching requirement. Matching was fluid and it affected the way we went about exploring the system although we didn't have a particular 'approach' and were just generally exploring it. It didn't lead to turn-taking.

Participant R The system could be a game. It could be described as a detection system with music but it was more like a game.

I seldom play music with other people. At the beginning I was reminded of my street dancing classes. I was trying to do something done by the other participant but then I thought of something of my own to try and I started to move something. I particularly liked it when we were trying to clap and heard something like a 'bling bling' sound.

The matching requirement sometimes affected how we explored the system. Our use of the system was simultaneous in the way two things led together.

Second interview

Participant L I preferred the second time. It was nicer and definitely had me thinking more musically. It felt more musical because I would look out for what he was doing and then think of sounds to go with his. I could hear his sensing and then mix it off with a couple of moves, like the rhythm, say. The interesting part of the second was how many different things you could do and making them work together. Maybe R would do the bell thing which would make me do the pad kind of sound. All these combinations made me carry on using the system.

I would not describe either time as difficult or challenging but the second was much more interesting and fun than the other one where you were more forced to cooperate. Here, we were looking out for what each other's doing which felt more like real cooperation. The matching prevents us from enjoying ourselves as much because we have to both make the same sound. It makes us both have to be part of a motion. The second session was more about the sound whereas the other way was more about the motions. I lost track of the time more in the second and was thinking less of what we were doing. Although we knew the gestures from the first session, I do not think the order we used the systems had much effect. If I were to use one system again, it would be the second one.

Participant R I preferred the second one. I had some kind of harmony that I really liked. It was fun and would be my choice if I were to use one system again. I would describe it as about the same level of challenge or difficulty but the first one was more stressful. When I was exploring it, I would have to think of something that I want to do. In the second one, I did not even think about my movement. I just thought 'I want this music' and did it. Exploration was guided by music rather than movement. Whilst I was using the second one, I almost forgot the time when I was trying to do something to explore the music.

C.3 Participant pair 3

The mapping was not explained to the participants.

C.3.1 Observations

First session

The participants got involved in what they were doing fairly quickly. They had a good rapport but they did not establish eye contact. There seemed to be a desire to explore the string mapping but their ability to do so was limited. They appeared to be trying to work together but found it difficult to communicate their intentions. There did not seem to be a single person leading. At one point both participants became quite slow and went into a more concentrated exploration. The session was stopped after 20 minutes by R.

Second session

The participants occasionally fell back into matching each other. They had a fair amount of rhythmic consistency and built up an intense sound by continually triggering drum loops. At one stage they took on roles with R doing rhythm and L the strings. At another, they swapped sides. Around 10 minutes in R slowed down attempting to explore slow movements but was having little effect as L was still making more active movements. L subsequently slowed down. 14 minutes in, R left the space and left L to carry on alone. 2 minutes later L leaves the space and R moves in to use it alone. However, the system has lost R's tracking data and is actually interpreting random input from his movement (although R does not realise this). After 3 minutes of this, R leaves the scene and the participants agree to stop the session. The total length was 19 minutes. There is a calm atmosphere at the end.

C.3.2 Reconstructed summary

First interview

Participant L The system was really amazing, for personal reasons. It's like a mind-body movement, virtual theremin-synth system - a pad that is triggered in a three-dimensional area rather than a two-dimensional area. I had to find out what it does as with any piece of software and getting better takes practice as it would with an instrument. Key to understanding it were finding the sweet spots and mechanics of the hand mapping. You have to go through the recognition of all of the sounds.

While using it, I was not trying to move more quickly or slowly but go with the changes on the rhythm. I wanted to be able to drag around the pad synth where sounds could be found and then moved somewhere. R mentioned incorporating layers in the system and this would be nice to work with the pad. You could have the same modulations of going forwards and sideways but then have three or four layers that you could play, perhaps lower, middle, etc. which would give you enough space. Also, you could have one hand on the pad and then the other on the drum loops or on the melody. But if it's a performance art piece then there would not necessarily need to be a melody or harmony.

The drum loops are OK but are stuck with a single tempo which you have to recognise. This is a pro and a con. Any loops sets a tempo and it would have been nice to be able to modulate this.

I though the whole thing lasted eight minutes. It would have been longer and flowed in a different way without the need of two people to be synced. This means you need a feather threshold area and it had an impact on the continuity of the experience. Without the matching requirement, I think we would have found our roles. It would also be possible to split the area into individual instruments (although this goes against the whole system). It would involve a different kind of trigger and you could give one person a tap tempo though I imagine this would be hard to do.

A key aspect of this is the sound it makes when you're not mirroring, the 'err' sound. It makes you question whether you are following or being followed. The difference between following and leading can become hard to tell. There are always going to be some microscopic differences between the participants unless you had about a month of preparation.

Participant R It was an unusual and abstract experience, like being a baby again trying to rediscover your environment. You're trying to find out where everything is, sync with the system and find the sweet spots. Early on it is quite disorientating because as a human you are visually oriented and you try to find some kind of visual target or visualisation of the space. You feel blind as you fumble around unsure of whether it is you or the other person triggering the sounds. It is a kind of reward system, or a weird dialect, sometimes it speaks to you, sometimes it does not.

There's something primordial about the experience, something within you that marks your territory, possibly ego or something more primordial that makes leading and following seem like a master and slave thing. You want to go back to a particular place and are trying to lead the other person there. The experience is not a competition (although maybe subconsciously it is), though I found myself intellectualising the process rather than going with it, which made it seem more competitive. I think we were always subconsciously questioning who was leading.

We would have used it a lot longer if we didn't have to mirror each other. It was quite physically exhausting due to the pace of the other person. I wanted to do my own sort of exploration and take my time, potter a bit, which I think anyone does when faced with a hands-on kind of challenge. Without the matching, I think we would have found our roles, for example the rhythm-oriented person, the synth-oriented person, ending up something more like Kraftwerk. We're speaking as musicians though; different participants would have a different opinion. Mirroring had a big impact on whether I could get into the flow. It made me more reliant. L was quick at the beginning and I was trying to keep up with him. I wanted to take the lead and explore a particular place but could not and felt like I could not express myself. I found it especially confusing being left-handed - I got confused between doing the mirror image or the mirror image of the mirror image. Our ability to use the system improved slowly - give it another hour perhaps.

The system would have been more intuitive as a solo performance. I think any kind of shared space imposes some kind of matching requirement meaning that you have to know the other person, intrinsically understand them. At the very least you need eye contact, which we did not have. With a more symbiotic relationship we would have been able to notice the more peripheral and microscopic movements. Moving slower would have made a more stable performance with sounds that were more coherently ordered and repeatable. Without more awareness of the other participant, the mirroring is more rough and the things that you find together cannot necessarily be found again. After you have discovered the sweet spots, then you can start moving a bit more quickly and making big sweeping gestures.

The mirroring was a restriction, which is not necessarily a negative thing, but it requires some kind of coherent structure and I think some kind of preparation would have helped with this. I used to do drama and we did similar exercises. It would involve starting off facing each other with eye contact, then noticing microscopic movements. Then slow movements would begin as one person leads and then another, followed by larger movements. Leading and following happen naturally. It is more of a transference, which we did not get here. I was trying to grab eye contact a lot and slow it down to find and explore those individual sweet spots. I think with more eye contact we might have had a more symbiotic relationship with each other and with the space, which would have made a more interesting experience. We also lost some momentum from having started off so quickly.

However, we did have some moments of continuity and flow. Often, it was lost when you heard the deviation sound. At first the sound is quite nice but then you realise it's a negative sound and it throws you back. It makes you start panicking and you have to start moving your body. You question if it is you or your partner. You tend not to be able to find that spot you were at before but return to something familiar, like that nice tom sound on the floor or the synth which gets you back into the flow.

I wanted to explore more of the system. There were repetitions of some sounds. Some I wanted to hear again but I wanted a broader palette as well. I was looking for something other than that synth, which just seemed to come there, swell and repeat two chords. I wanted to find other sounds and I think that there might be some that we did not find. We kept finding that drum down there on the floor too. I wondered what else was on the floor and wanted to go down and have a look but could not. Some sounds seemed to be triggered at the same time and they might have been separated better.

The synth modulates outwards when we come here and became quite distorted when we went there. I was not sure if it was a sample or had a specific rhythm but we were clearly triggering a specific loop. I would have been interested to change the tempo of the loop and modulate other sounds rather than just the pad. It might be good have the system in layers.

The system could be set up without the mirroring and then you could have, say, one side of the room does percussion, one does everything else. That would be interesting (and scary). The system was restrictive in having a fixed tempo which is understandable for the samples. I guess a tap tempo might be quite difficult as it is getting quite advanced. Losing the mirroring might go against the whole system, but the system is meant to be broken, experimented on. You could have an invisible band with a guitarist and a virtual theremin, though with different technology as a theremin would probably be using microwaves.

All this said, I am a musician and so would probably express myself and have a different opinion than, say, a group of expert dancers. The system would be interesting if used by interpretive dancers and used as a dance oriented detection thing. They would probably give it a much more positive response, treating it less as a sound thing and being more about the space it creates. They would probably use it for a lot longer and might just naturally resonate better with the system. You could do some yoga experts on it.

Or it may suit performance artists and be a part of something like a Jean-Michel Jarre performance. As a performance art piece, we would have much more abstract and less rigid expectations than we do coming from a musician's environment. As a movement piece, our bodies themselves would be something to be explored. It is still relevant for musicians though. It is a bit like a Reactable. It would be cool with a few of those ideas involved, like being able to grab a modulation from there, activate it and then drag it over here.

Second interview

Participant L This was an improvement on the first one. The matching thing was interesting but this one definitely had more control and was much more fluid. But it would be confusing as a first experience if you do not know what is being done by each part of the body or which sounds were being made by you. Who was making each part was a curious aspect which we found out through repetition. Learning to use the system through discovery while using it did not get in the way of having fun because, whether you are controlling it or experimenting, it is all about the flow of your body movement. The sound was always coherent, in time, with harmony, with rhythm, with tempo. It was like we were walking on the same path, in the sense that it was very safe, like a road with flowers and trees and each one of us goes for one thing and that works out just fine.

Having multiple instruments to control makes it more difficult to focus on one sound. We had random exploration, where you just go through it, which was followed by more technical aspects.

The triggering of the samples is not absolute. For example, with that kick the velocity does not seem to respond to speed the same way each time, which you would expect it to. It was hard to realise how it worked. But sensors are easy to trick. You always wonder 'is it here?' You have a constant doubt about whether there are any uncontrolled external factors making the sound, which can be confusing. It reminds me of a book called 'How real's real?' which talks of a study where people have to hit three of eight buttons and then a central button, which displays whether the correct buttons were pressed. The participants have to figure out what the next sequence is but in the end it turns out to be absolutely random. The study investigates how the human brain always has a need to find coherence in a system and I think people struggle in the same way with this system. This doubt about coherence is the biggest struggle for when you are just starting although it does not stop you from getting into the flow.

I would describe the system as challenging in a way similar to when you play with someone for the first time. But the purpose of the system makes a big difference on your response to it. It could be an experimental performing interactive art piece, like those that are on tube station steps. People go around and marvel themselves for ten seconds and move along. That is a different thing from if it were an instrument, like the Reactable where you play with it and marvel and the tangible pieces and make some wild sounds. An instrument you would get along with after some hours of training - a guitarist does not know where the notes are when they first start. But the Reactable is meant to be experimental like this system. Even when played on by professional musicians it has a lot of randomness. There will be variables, mistakes and differences from the exact way you wanted it to play so even though it has visual content and the performer knows what is going on there will be happy accidents. But on this system, however much control you add, it will never be an absolutely accurate instrument and always be very experimental. You could perhaps solve a performer's need to know where everything is with something simple, like a virtual rack or a net of intervals made out of thread.

It is a bit of a stupid thing but the purpose of the mirroring one could also be for one of those team-building things, in the sense of working together with someone because it requires joint effort. Only, you could have a much more dramatic and really disturbing deviation sound. What would also be good would be instead of being its own sound, it was reflected within the sound as dissonance when you were not matching.

Participant R The second time round was an improvement on the first session. It was much more fluid, intuitive, enjoyable and fun. It was more interesting, sonically, as a performance and just to look at. I felt more free. In the first one I wanted to stop but this time I was more happy to continue because I was getting more out of it and still finding sounds. I am glad I had the previous experience though as it made me appreciate this experience more. They were both challenging but in different ways. The first time because you were playing blind. That was frustrating because you wanted to get it right. The second time the challenge was on a different level. You wanted to have a more semantic exploration and find these shapes. Being less bound we explored more. I am still not visually there in my head but I am more of an abstract person than visual.

How you explore it is like when you do photoshop or learn a new piece of music software. You do not read the manual or calibrate everything but just go mental and start pushing buttons and learn it through self-discovery. You find a sticking point, your anchor. From there you find maybe another anchor that you are familiar with and you can move between these kind of like call and response and lets you find out what is instrument, what is sound. Moving between the two leads onto the kinetic aspect, which means it is about moving as well as watching. This brings in more of the sounds, more exciting sounds and might lead you to create a more active sound. When there is a lull you return to your anchor.

This type of learning is fun though I guess that is not a necessity. It does not necessarily lead to learning the whole thing and it might not be so good if the system was to be used as a technical tool or a performance related thing but as a learning tool, a novice performance education tool this would be fun to have. The age of the user would probably affect how they respond to this approach with younger people responding better. If it is an overall tool for just everybody then children will find something from it. Only they would be exploring by running round, having fun and much more intuitive. This random exploration does eventually need to be followed by the technical aspect which we did though we were pretty familiar with this from the earlier experience. I think for more exploration we would need more control. I suppose it is in the innate nature of human beings to want control - particularly men!

But your response will depend on your background. We are from a musician or performance background so are talking in terms of what we like to see technically. We have preconceived notions from things we have seen on Youtube like the Reactable. As a performance art kind of thing, I guess something like this has already been done but maybe not as an instrument or a virtual band. But a guitarist knows where his guitar is. A drummer knows where his drums are. It could be like a quirky Reactable kind of thing. It seems like a natural progression from needing a physical table and objects to modulate and create the sound. As an instrument, we could have a space here with a virtual rack of effects, etc., which could be done with markings. It would be interesting to know what the system is for and how far in the development it is, though not make it any more or less fun. The calibration will be broad and depend on how it is used, which will depend on your expectations of what it is for and what it can do. A performer would need to know where everything is. Someone more technical would also want more control, to be able to calibrate it to make very precise movements and place sounds where they like. Performance artists would maybe want to be able to have it more random and discover the system with the use of their body.

Repetition is not always successful. This can be fun as it adds to the glitch element, which is quite cool. But I wondered whether it was perhaps a certain movement or certain speed it was responding to because there were more sounds when I was making much bigger or faster movements. It may have been that the sweet spots were not necessarily static. Where everything is did not feel static. I stepped out of the performance space because it seemed to have come to its conclusion but this allowed me to observe L. This was not planned. When you start observing one person on their own you start seeing that there definitely is that sound there. It let me improve my understanding of where everything is and made me want to know more. So after watching for a bit I returned to the space and L came out.

But a bunch of sounds do not seem to stay in the same place. There was that kind of staticky sound here and that modulation there going up and down that seemed like it was operating some kind of wave or cutting through a stream. I was not sure where the percussive things were. They seem to be randomly placed rather than more specifically like in percussive groups. That tom down here misses sometimes and did not always trigger, which would make you think 'oh shit. Down there it seemed more resonant and further left more verby though that might have been because of that kind of static on the other side. It was not always in time which is something you would want it to be. It was possibly synced to a sixteenth or a thirty-second. The drum loop I could not find. It seemed to stop when you got slower and stay in the more kinetic you were. Sometimes it seemed more an amalgamation of sounds if you move a certain speed or a certain direction rather than individual sounds that you triggered. I think having multiple instruments to control leads you to rationalise it and make it more logical. It is human nature to look for coherence and solutions to problems so when you doubt this you start the repeat thing. It does not stop you from getting into the flow but it is a struggle if you have the study L mentioned in the back of your mind!

Overall though, the system felt very safe. Whatever we did, it was going to produce a coherent thing, in tune that would be some kind of loop if you laid it out on a computer. This second time it had less or hardly any of the 'uhh-uhh' sound. I think it would be better without this sound - deinhibiting and less intrusive to the performer. I can see you might need it if you were going for the motivational team-building thing but if that was its purpose I would take the drama route where they teach you to look in the eyes, do the movements slowly, build, start from that, and so on. I suppose this is programming the other person to behave in a certain way. But we did not have this introduction here so the matching was more experimental. The system worked better without the matching but I think I am too selfish for it! I suppose you get used to it, like breathing underwater, but I found we were much more symbiotic without it. We still found ourselves finding specific roles. However, there is more doubt about what they are generating and what you are generating, whether it is just an amalgamation of sounds based on all of your movement, what the other participant's true reading on the sensors is. That is why I wanted to swap and explore the other side. This allowed me to get more spatial awareness. Even then it felt like there was more to find. It would be interesting if the sounds from each person were played through different speakers. Or it could be the opposite, so my speaker plays his sounds and vice versa. It would be very interesting to have each sound coming through its own speaker. As a solo performance you would become more in tune and confident.

I would love to have something like this where you can use your own rhythm. I think yo use it daily would be quite inspiring actually. I always admire artists like Aphex Twin and Squarepusher who do erratic jazz timings and all but I get really frustrated that I cannot sit at home on music software and program something as random and chaotic as that. I love random elements in my own music and would love to have something really glitchy like that. I have a theremin at home. I do not use it very much as my setup is crap but I would like to. It is not a proper Moog theremin which is a proper thing with a proper box. I would probably use it more if it was. But as I am sure you know, the theremin is not played by touching but by playing with space and that appeals to me. When I was younger I loved movement and things as well as dancing and drama.

C.4 Participant pair 4

The mapping was not explained to the participants.

C.4.1 Observations

First session

The participants quickly established a good rapport and were laughing and communicating successfully from the outset. L was quick to try extremes and led the participants towards interacting on their knees, facing backwards, jumping. However, throughout the participants discussed their interaction as a joint activity. At one point L stood on his head, at which point the investigator interrupted and asked him not to do so as it was likely in breach of the study's risk assessment. The skeleton tracking did not work well in this session, sometimes able to track only the torso. The participants asked to stop after 16 minutes without a clear leader.

Second session

There was again a lot of raucous movement and the investigator asked the participants to be careful not to hurt themselves. They asked to stop the session after 6 minutes, again without a clear leader.

C.4.2 Reconstructed summary

First interview

Participant L At the beginning it felt like a learning experience to get it to work at its best. We were going between static points exactly matching. We had a strong sense of not wanting to make the 'nn nn nn' sound from not matching. The sound is not horrible; it is actually quite a nice noise but you are still aiming not to hear it. Sometimes, the system still made the noise even though we were matching. Towards the end, it was more freeform with us trying less hard to match. I was much more immersed in the experience at the end and the time did not seem to last very long. Generally, I was less conscious of what was going on in the fun parts, for example when R was on his head but it still seemed to be working. These bits were just funny really. It felt like we were aligned with each other and I forgot about what we were doing with the system.

I have only really said 'hello' to R before and I felt more self-conscious about working with someone I did not know than I did about moving my body. That said, whilst I do dance, I probably would not have signed up if someone had asked me to just come and dance. If it was just me moving, I probably would have felt really self-conscious (without alcohol at least). I did not feel self-conscious about the camera. But I was not really dancing as I felt I needed to know what was being done by the system. I would not say we were talking or problem solving so much. It was more just movement and it became more fluid as the session went on. I do not think the system would be very good for dancers as there is no pause in the sound so there are no moments to stop and regroup. When you are dancing, there is a shift between the songs or the tracks of the music.

That we had to match each other made me think initially of leader and follower, which is established quite quickly. We did not have very much eye contact but were more just able to see what each other was doing with our peripheral vision which was enough. It was quite bonding between us which was nice. The matching is quite an intimate thing but in a friendly, playful environment which made it feel very safe. Without the matching, I think we may have found the sweet spots and triggers more quickly. We would just be zipping around as there is no negotiation. So we would probably use it for less time unless we started unravelling things that we were not able to get to this time. But it would not have been as much fun on your own as the shared experience is a big part it.

We both had quite a similar approach to exploring the system. We had small strategies to find one thing. If that worked then we would try moving around. After trying something in a different context, we would return to something we knew worked. We were quite constrained in what we did. We remained enclosed within the markings and did not go out of them at all. Perhaps we would have done if there was more of a gap around. You cannot run around in this room.

I think it would evolve the more you used the system, if you had the same partner.

But generally, there seemed to be quite a limited set of sounds. There may be more to it but we kept hearing the same stuff again and again, which made me feel that there was not much more. Whilst the 'voom' sound was really responsive and had quite a gross movement, there was not so much fine 'ding-a-ling-ding'. The feedback was not great. I wanted to control the sound more and was trying to find a way to make the music, rather than have it react.

Participant R The experience was cool - slightly strange but really enjoyable. It was slightly stilted in some way. It became more fluid but we were working it out rather than dancing. We were quickly problem solving to get it working and at the beginning it felt more like a goal. We were going to test the system and while doing that find out its limitations. To start with, I was thinking of shapes rather than movement and it became less about the dynamic and much more about static points, controlling the system from one point and moving from form to form while keeping in the same position. Even if there was a small amount of movement, it was about keeping that movement the same and avoiding the deviation sound. However, the system was working at some points even though we were not in the same position. At other points even when we were pretty much exactly matching bar a finger it would not work.

I realised for quite a bit of it I was not looking at L. I was thinking less explicitly of what L was doing, relying more on feeling a sense of it through peripheral vision and then occasionally checking how I was doing. It felt nice to get sorted. We had little bits of talking, which was good, though not right at the beginning. Leading and following was quite a bit in my mind. I do not really know L at all and neither of us have ever done anything like this but we had a playful relationship and the experience was a really fun way to get to know her. You should do it with everyone!

I do not feel self-conscious about movement, less so than with public speaking for example, and I forgot about the camera. But it would not have felt doable if we were just asked to do some dancing. It gets you moving and felt doable because we were expecting a controller. That said, we were restricted compared to what we could have done. It was pretty much rotation and extension of the arm and movements that the system led you to do. You could not include things like compression and opening stuff. We did not even touch going beyond the tape. But it felt weird moving in this room. In a way it was a bit like a dance floor but it makes a weird space for movement. I think in a different space we would have explored beyond the marked area.

It was more just movement than dance. For dance, there are more dynamics in the music and it tends to be building but the sound will never do this if it is always following your moves. I found sometimes you want to react to the system as opposed to it reacting to you but you will always be chasing your tail if you try this.

We were initially made hesitant and unsure by the matching requirement. It felt kind of restricted in some ways but it got more dynamic. It made it easier to learn to use the system and I think it would definitely have been shorter without it as you would try lots of things quickly. Possibly it prevented us from unravelling the possibilities but on the other hand it might make you lead each other into the sweet spots and find more of them.

We definitely had a number of exploration strategies but I could not say what they were. When we found a little thing - like when we were going 'voom voom' - and it worked then we would try to break that up and find a new thing. Maybe if that worked then we tried something totally different, or maybe the same thing in a slightly different context, for example on the floor or with our legs. At times I felt like jumping across and doing slightly more silly things, like standing on my head.

If we were to use the system again, we would get more refined and with the same partner it would evolve. I would maybe try different things or get a little bit more exacting with it. However, if I had to come every day for a week and it was in the same space, it would get tedious. A different space would lead to a different experience. For example, I would have found out what happened beyond the tape somewhere else. But I don't have a sense of really wanting to explore this more. It did not feel like it would lead to more discoveries. To explore this, I would want to try it in a different context. That would make my mind start working, asking 'How do I use this? And what for?' It would make the repetitive sound less of an issue. But I do not feel like there are more possibilities to explore. Or maybe there are but I do not feel motivated to explore them. It was not like a program where you might think 'I can't do very much' but there are so many possibilities. I think the feedback was a bit 'blehh'. I did not feel like I really worked out what was going on. I did not need to know everything but the amount of movement and exploration we did justified more.

You should be able to have more musical control. It felt too much like you were cutting or fading into a piece of music that was already playing. The bit where we were going 'voom voom' with the synth was less like that. That was actually something that makes you think 'yeah that's good'. Some parts were fun. Towards the end it was great and I definitely lost track of time a bit. I think the whole thing lasted about 20 minutes.

Second interview

Participant L The first system was a dynamic movement matching system that responded sonically to how well you mirrored the other person. I am not sure if I agree with R's expression 'forced collaboration'. Perhaps 'mutual collaboration', 'compulsory collaboration', or something else.

This second session was less tight in the shared experience and less focused than the first one. I preferred doing it with someone else. That was more of an important factor in how enjoyable it was than how careful we were. The togetherness and having to match meant there was more of an impetus to learn. The two systems gave a slightly different sense of achievement but the first one had more. The deviation sound, which was not there in the second, put in place a kind of reward system an gave a sense of a goal.

R has some dance experience but I am not a dancer. Towards the end of the first session I was feeling less self-conscious. But in the second session I was more unsure of what to do next and when I was trying to stop, it was because of this, as well as the repetitiveness of the music.

The system seemed less responsive in the second session. The sound went in and out and there seemed to be less attention on me. It may have been that the responsiveness was less obvious in this session because we did not start with small movements like we did in the first. I think this time ended up being mainly about the dynamics whereas the first one was about the actual shapes you are forming as well.

Participant R The second session I was a bit more just dancing around. It was more us just moving, which was nice. Our movements were more tied to the dynamics. It had less of a goal and was more open from the start. Perhaps 'goal' and 'open' are not the right words. It was definitely less focused. There were points of mutual influence and moments when we reflected each other but it felt more like we were just both engaging in the same experience rather than a shared experience. We were still aware of each other but mainly just of where in the space the other person was whereas in the first one it was also about having the same shape. Perhaps if we knew each other we would have had more awareness of each other and start interacting and doing stuff like a new composition. I do capoeira a lot but have not been much of a dancer recently. If we were improvising dance or doing contact improvisation, we might start by just flinging around but eventually stop and ask which bits worked and then try them again.

It might have been different if it was in a different space or perhaps if we had longer, although I guess time was not a factor as you did not set it. But the first one seemed to be about us interesting ourselves in each other movement-wise as well as finding out what was going on. We were much more considered and careful in our exploration and curious about what would happen. It was fun to think of what to do next and felt like there was more control. When we started the second one I wanted to figure everything out but to be honest pretty soon I was just not as bothered. By the end I was just looking to see if the system was going to do anything. It felt more like a soundtrack. The tempo was the same and although there were more dynamics than the first, it was like we were just catching to the music. Though when we both stopped and I made a sudden movement, it did respond with a 'DUHH RUHH RAHH!'. At the end of the first one I felt less unsure of what to do than in the second, and we had both become less self-conscious.

I would describe the first system as a responsive audio system with forced collaboration. It is difficult to put it into a sentence though. The collaboration felt forced because it will not work unless you collaborate with another participant but it was a more positive step. 'Forced' is a bit of a negative word and but that aspect was not necessarily bad. 'Compulsory collaboration' sounds even worse though! The first one had more of a build up which happened when we were just fooling around. I think the two systems serve different purposes but they are both fun in different ways. Fun is quite subjective term as it can be learning and controlling and eating something or just throwing yourself around.

C.5 Participant pair 5

The mapping was explained to the participants.

C.5.1 Observations

These participants were exceptional within the study as they were friends and somewhat older than many the other participants. They seemed less confident that others about using the system before starting and agreed that R would be 'leader'.

First session

The participants began by performing familiar dance moves, matching in style though not in actual pose. R was actively leading. Their movements seemed quite unrelated to what they were hearing. At 9 minutes in they began talking to each other quietly. At 14 minutes in, they were conversing about matters completely unrelated to the study, still moving quite vigorously. They then started whispering and began matching without reflecting each other. The session was ended by me after 20 minutes. They were quite relieved that it had ended. Although it was quite clear that it was up to them to decide when to end the study, they seemed to have carried on out of a sense of obligation.

Second session

Prior to this session it was restated that the participants were free to use the system for as long as they pleased. L asked if the investigator would prefer that they use it until he stopped them and he again said that that was not the case. 10 minutes into this session, the participants tried standing still for the first time. The session was stopped after 11 minutes, led by L.

C.5.2 Reconstructed summary

First interview

Participant L It was an unusual experience. We caught onto the fact that it was our movements that were causing the sound. I do not do this kind of structured movement. We were both playing on the kind of movement in exercise, which is more structured. We were putting our arms and fingers in the air and liked the sound. R may think differently though as we did not discuss what we were going to do beforehand. We came to recognise the sounds and both found together movements that we quite liked the sounds of so tried to recreate them, which I suppose is just psychological. Sometimes the same movements did not create the same sounds though perhaps we were not in unison with it.

The sounds we created while we were doing it reminded us of things from the sixties, the hippy days and we were playing on that. We tried to be rhythmic to see what would happen but it was not really rhythmic. It was more limited too because we had to keep within that space, which prevented us from doing certain dances. Not that we're very proficient at that. I have no coordination and cannot really do this in time. I do not do that sort of group thing either because I do not like it. We are not the sort of people out in clubs doing dances that people do today. I do not know what they do today but I suppose they might automatically follow people. By the end, we were getting bored and trying anything.

We are friends and definitely had a clear sense of who was leading throughout. I did not think I was copying R very well. Copying each other is difficult when you do not know what you are doing. I probably would have got lost in it because you can go almost into a trance but I could not because I had to be aware of what R was doing. Or maybe I would have just pretended to be lost in it. I had had enough really. I do not know how long the session was. It felt like more than ten minutes but was probably less than ten.

Participant R I love music and movement and that attracted me to the study although I had no idea what we were going to do here. It was hard. We were asked to move around and as you moved the music or the sounds changed and so we tried to be more inventive with our movement to create the different sounds.

I quite like just moving about. Copying each other was OK because I was leading which, logically, is easier. I was impressed by my friend, L, who must have found it hard to copy my movements. She was very good at doing this even though she claims to have no coordination and said she could not but it is just movement. I was first of all trying to recognise movements that I'm familiar with. We were playing on exercise movements and tried to do movements that made nice sounds.

The way we used it changed over the session. Initially, I was trying to find the system's possible sounds. I was trying first of all to recognise movements I'm familiar with. We did quite a lot of trying to explore what was possible but all within the perimeters. I said let's do a dance thing, let's do something from class, throw a tantrum, what kids do - anything, just within that space. It was quite spontaneous. I do not like group things like this apart from the Zumba. But I thought I was repeating myself and asked L to take over. It was OK but quite difficult to copy L - difficult without a mirror anyway.

It might have been me being a little bit out but I was expecting from what you said that if you did one thing and the system did something then if you did it again the system would do the same thing again. We had quite a lot of raising the arms and stuff like that which let you sense changes in the different sounds but they didn't always seem to follow your movement.

The sound reminded me of music from the sixties. It was maybe rhythmic and we ad libbed as we went on. I probably could have got lost in it but I did not because I had had enough really. It got a bit monotonous towards the end. L kept asking how long it was going to be and when we would be out of here - it was like going out with my daughter! I do not know how long we were doing it for - maybe 15 minutes. I have never seen anyone else use the system. It must be fascinating for you running this study to see all these people use it.

Second interview

Participant L It is quite difficult to draw a preference between the two systems. I would not say we explored them in different ways. If I had to recommend one, I would choose the second. It was easier because we were doing our own thing whereas in the first one I was copying R at first and then R copying me. We could have been given some time to make a plan but I think that would probably put people off doing it. I thought R would prefer the other so I am surprised that she would also recommend the second. I think the second one lasted more than five but less than ten minutes - maybe six minutes. I looked at my watch in this one and I was much more aware of the time. It went slower perhaps because you did not have the experience of watching somebody else. In the first one I wanted to stop earlier but we thought we should not because we have come here to do the thing. In the second one though I was thinking 'Oh God I want to stop' because you are standing there doing all these movements.

The second one did not seem to create as much noise reaction. The sound was not as powerful except perhaps the hand thing. That made me do the hand thing more because it felt like it was brining more sound than the lower part of my body. The nice bits were when we were doing it together because the sound was better.

I think we did not behave differently because we were creating the noise ourselves. Perhaps we might have done if we were not confined to this space.

One movement made a sound similar to what you hear at the spa. I thought I was a city person and into loud noises and stuff but then I went to that meditation day down the road. It involved just sitting on my knees doing what you are supposed to do. I could not believe how quiet I was for the whole day. I was entranced by it. It was weird because I did not think 'Oh my God I must use the phone' as I would have excepted to.

Participant R If I were to recommend one system to someone I would choose the second one because you can do your own thing, the same as L. L probably thinks I would choose the first one but the reason I enjoyed that one was because she had to copy me which meant I was able to do my own thing. We decided that L would copy me and then I would copy her. So I did also copy her as well. I do not think I have asked someone to copy me since I was a little girl. Inventing things to do was made harder knowing that somebody had to copy me though and I could not have carried on doing that. Doing your own thing is nice. That said though, I think the first one was more fun because it was completely new and we did not know what to expect.

There was not much difference between the sessions in the way we explored the systems. The second one I started out trying to figure out how my movements made the music change. I tried to see what the noises were but only really recognised the sort of raised hand thing. I found it hard to decipher. I would think I had it and then found that I did not and that was annoying. If I did something like that then I thought 'aha that's it' and then I started it again and it did not do it right away. It was very slow so maybe this was because my movements were quicker than it. I like to move quicker than the system was though. This was annoying me, maybe because I am older so I know what I like to listen to and I expect to know what the next thing coming is going to be. Towards the end it had some nice noises but I wanted to stop because of my cigarette.

There are only so many types of movement you can do to this type of music and I think it limited me. The music makes you tend to do samey movements. It is not something I would have chosen to play for moving to. I was responding to the music and I would behave differently if you had put Michael Jackson or the Beatles or something on. I did not move as quickly as I like because I have to wait to be inspired by the beat. It is not like class because you have to do your own thing. The music reminds me of the stuff you hear at a spa. We go to a spa abroad. The music is meant to be soothing and everything but it irritates me after a while. I am not the sort of person who goes for that kind of stuff. I am more of a city person. I like loud noises and stuff.

I am an artist though and I imagine this kind of music could conjure up images and abstracts and stuff for painting. Art and music might be interesting - we have done that before. Maybe that could be your next PhD!

C.6 Participant pair 6

The mapping was not explained to the participants.

C.6.1 Observations

First session

The participants instantly established eye contact and a good rapport and were quickly laughing. However, there were technical difficulties, and the skeleton tracking continually lost its calibration on R. She recalibrated a few times, and the system was restarted to try to resolve the issue but the difficulties arose again. R was asked to change into trousers in an attempt to resolve the issues. The session was restarted (9 minutes after the original beginning). The participants were asked not to make any contact as this had previously been found to cause difficulties. The system began responding slightly better. There were a number more recalibrations and at one point the participants were asked to turn more towards the depth camera. The participants asked to stop the session after 19 minutes (counting from the point it was restarted).

Second session

In the second session, the technical issues were less severe. At 6 minutes, R demonstrated the percussive stab mapping to L. The participants asked to stop the session after 8 minutes, possibly led by R.

C.6.2 Reconstructed summary

First interview

Participant L The system is interactive and did not always work OK. Sounds are made according to our movements. To use the system I had to work with another partner and to make it sound good we had to have identical movements. Mirroring meant that we can move together which made me feel more comfortable. We always had contact and communication happens all the way so we were aware of what was going to happen next. Sometimes I led and sometimes R led. I would not really know what to do if asked to use it individually. I am not a dancer and not so good at creative movement

The beginning was just trying things out. Towards the end it was more familiar and we knew how to fix stuff so we felt more comfortable and confident. I think we were using it for around 30 minutes. I am surprised that R thought it was only five minutes. It was definitely longer than that - at least 15-20 minutes. We stopped in the end because we did not know any more poses that can cause other effects and have an impact on the sound.

I am a musician and play with others. I play the drums and the bass guitar a bit. I think this is quite different from when you play in a band. You cannot all do identical things in a band. The most similar aspect is the communication. But when I play with others I focus about 60 per cent on myself and how I organise my parts and can maybe focus 40 per cent on the others. This felt like it involved a 90 per cent focus on the other person and about 10 per cent on what I am going to do next.

Participant R The system is fine. It has an interactive soundscape thing where your movements are mapped. How far we both mirror each other and maybe how fast and repetitive we are triggers and creates sections of audio. It is OK.

I think we were using it for about five minutes. I am surprised by how long L thinks we were using it for. Maybe it was seven or 10 minutes. I was a little bit lost in it - just looking at each other going 'woo'. I was thinking about copying, watching each other's movements and listening to the sound. We found it interesting but stopped when we did not really know what else to do and also my arms were starting to hurt.

Doing the same thing together made it feel less like I did not know what I was doing which made it less scary - especially with us being recorded. It meant we were in it together and it was not just a single thing. We took it in turns as to who was leading. I think it would be more difficult if we were not doing the same thing. It would be a bit scary and I would not know what to do. It would also be more difficult to learn how we were creating a particular sound as we would not know where our sound was coming from and whether it was us creating a particular sound. But, like L, I am not a dancer. Maybe an actual dancer would not find it difficult without working with someone else as they could do all this creative expressionist stuff.

Exploring the system involved seeing if we could find out what our movement did to the sound. It was half and half between finding all of the possibilities and focusing on one small part. It was mainly just finding what sounded good but then when we found a sound we liked we kept on making it. We found that we could modulate the sounds, like a 'wowowow' sound. It became more intense if you go up a bit and the more you do it, and the more repetitive, the more there is and the track starts to play.

I play the guitar. I used to play in a band but not so much any more. I am more of a techie. I think this experience is different from making music. If you are making practice sounds then you know when other people will come in. It might be more similar if you were improvising but I have never done anything purely improvisational.

Second interview

Participant L The second session was a bit shorter than the first one (if you do not include the time wasted at the beginning when it was not working). If I were to recommend one system I would choose the first one. I have no clue how the sound works in the second one. It seemed more random - like pure sound effects coming together without any systematic events. In contrast, the first one felt like it had a beat. We found ourselves moving rhythmically and the music came up with a rhythm. For a while we were really going on it. It was more fun, collaborative and interactive than the second.

I did feel a bit safer in the second one because when we were doing identical things I had to consider whether R would be able to match what I am doing. This meant I tried more things that were different.

Even though we were not mirroring, I still wanted to have a relation with the other partner in the second one. It does not make sense to do our own things without some kind of correspondence. But it is more likely to be trying to cause her to do the opposite thing from me rather than matching.

Participant R If I were to recommend one system I would choose the first one. I did not find either challenging so to speak but I felt more comfortable mirroring. I found it difficult this second time round to know where the sounds are coming from. Doing our own thing meant things were not as simultaneous. It was confusing because it was difficult to know if it was me making a sound. I felt more like I did not know what I was doing which made me feel a bit more silly than before. I almost started mirroring L automatically.

In the second one it seemed like the sound was already playing continuously with us just having an effect on it whereas the first time it felt like we were actually making the sound.

We stopped when we did because it felt like we had explored as much as there was. I am not sure how long this second one was. I do not know if I am very good at estimating time but I would say it was definitely at least a bit shorter than the first one.

References

Abdallah, S. and M. D. Plumbley (2009). Information dynamics: Patterns of expectation and surprise in the perception of music. *Connection Science 21*(2), 89–117.

Antaki, C., M. Billig, D. Edwards, and J. Potter (2003). Discourse analysis means doing analysis: A critique of six analytic shortcomings. *Discourse Analysis Online 1*(1). `http://extra.shu.ac.uk/daol/articles/v1/n1/a1/antaki2002002-01.html`. Accessed 14 April 2012.

Antle, A., G. Corness, and M. Droumeva (2009). What the body knows: Exploring the benefits of embodied metaphors in hybrid physical digital environments. *Interacting with Computers 21*(1-2), 66–75.

Arfib, D., J. Couturier, L. Kessous, and V. Verfaille (2003). Strategies of mapping between gesture data and synthesis model parameters using perceptual spaces. *Organised Sound 7*(2), 127–144.

Ariza, C. (2009). The interrogator as critic: The Turing Test and the evaluation of generative music systems. *Computer Music Journal 33*(2), 48–70.

Arnstein, S. R. (1969). A ladder of citizen participation. *Journal of the American Institute of Planners 35*(4), 216–224. Cited by Jo and Tanaka (2009).

Auslander, P. (1999). *Liveness: Performance in a mediatized culture.* Abingdon, NY: Routledge.

Auslander, P. (2002). Live from cyberspace: or, I was sitting at my computer this guy appeared he thought I was a bot. *PAJ: A Journal of Performance and Art 24*(1), 16–21.

Auslander, P. (2005). At the listening post, or, do machines perform? *International Journal of Performance Arts and Digital Media 1*(1), 5–10.

Austen, K. (2010). Looking for art among the gadgets. *New Scientist CultureLab blog*. `http://www.newscientist.com/blogs/culturelab/2010/02/looking-for-art-among-the-gadgets.html`. Accessed 14 April 2012.

Aylward, R. and J. Paradiso (2006). Sensemble: A wireless, compact, multi-user sensor system for interactive dance. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, Paris, France, pp. 134–139.

Bailey, R. A. (2008). *Design of comparative experiments*. Cambridge University Press.

Baker, C. (2004). Membership categorization and interview accounts. In D. Silverman (Ed.), *Qualitative Research: Theory, Method and Practice* (Second ed.)., Chapter 9, pp. 162–176. London: Sage.

Bandura, A. (2001). Social cognitive theory: An agentic perspective. *Annual Review of Psychology 52*(1), 1–26.

Banister, P., E. Burman, I. Parker, M. Taylor, and C. Tindall (1994). *Qualitative Methods in Psychology*. Maidenhead, UK: Open University Press.

Baranes, A. and P.-Y. Oudeyer (2012). Active learning of inverse models with intrinsically motivated goal exploration in robots. *Robotics and Autonomous Systems in press*.

Barthet, M. and S. Dixon (2011). Ethnographic observations of musicologists at the British Library: Implications for music information retrieval. In *Proceedings of the International Society for Music Information Retrieval Conference*, Miami, FL.

Bayliss, A., J. Sheridan, and N. Villar (2005). New shapes on the dance floor: Influencing ambient sound and vision with computationally augmented poi. *International Journal of Performance Arts and Digital Media 1*(1), 67–82.

Benford, S. and G. Giannachi (2008). Temporal trajectories in shared interactive narratives. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Florence, Italy, pp. 73–82.

Benford, S., G. Giannachi, B. Koleva, and T. Rodden (2009). From interaction to trajectories: Designing coherent journeys through user experiences. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Boston, MA, pp. 709–718.

Benjamin, W. (2007). "Sweet anticipation: Music and the psychology of expectation" by David Huron (book review). *The Journal of Aesthetics and Art Criticism 65*(3), 333–335.

Berlyne, D. (1970). Novelty, complexity, and hedonic value. *Attention, Perception, & Psychophysics 8*(5), 279–286.

Berlyne, D. E. (1966). Curiosity and exploration. *Science 153*(3731), 25–33.

Bernhardt, D. and P. Robinson (2007). Detecting affect from non-stylised body motions. In A. Paiva, R. Prada, and R. Picard (Eds.), *Affective Computing and Intelligent Interaction*, Volume 4738, pp. 59–70. Berlin: Springer.

Bevilacqua, F., L. Naugle, and I. Valverde (2001). Virtual dance and music environment using motion capture. In *Proceedings of the IEEE-Multimedia Technology And Applications Conference*, Irvine, CA.

Bevilacqua, F., N. Schnell, and S. F. Alaoui (2011). Gesture capture: Paradigms in interactive music/dance systems. In G. Klein and S. Noeth (Eds.), *Emerging Bodies*, pp. 183–193. Bielefeld, Germany: Transcript Verlag.

Bilda, Z., B. Costello, and S. Amitani (2006). Collaborative analysis framework for evaluating interactive art experience. *CoDesign 2*(4), 225–238.

Blackwell, T. and P. Bentley (2002). Improvised music with swarms. In *Proceedings of the 2002 Congress on Evolutionary Computation*, Volume 2, pp. 1462–1467.

Blackwell, T. and M. Young (2004). Self-organised music. *Organised sound 9*(2), 123–136.

Blaine, T. (2006). New music for the masses. *Adobe Design Center, Think Tank Online*. `http://adobe.com/uk/designcenter/thinktank/ttap_music`. Accessed 14 April 2012.

Blaine, T. and S. Fels (2003). Contexts of collaborative musical experiences. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, Montreal, Canada, pp. 129–134.

Blaine, T. and T. Perkis (2000). The jam-o-drum interactive music system: a study in interaction design. In *Proceedings of the 3rd Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques*, New York, NY, pp. 165–173.

Blythe, M., P. Wright, J. McCarthy, and O. W. Bertelsen (2006). Theory and method for experience centered design. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (extended abstracts)*, Montreal, Canada, pp. 1691–1694.

Boden, M. A. (2005). Aesthetics and interactive art. *The ACM Conference on Creativity and Cognition*, London, UK. Keynote talk.

Bollo, A. and L. Dal Pozzolo (2005). Analysis of visitor behaviour inside the museum: An empirical study. In *Proceedings of the 8th International Conference on Arts and Cultural Management, Montreal.*

Bongers, B. (1994). The use of active tactile and force feedback in timbre controlling electronic instruments. In *Proceedings of the International Computer Music Conference*, pp. 171–171.

Bongers, B. (2000). Physical interfaces in the electronic arts. *Trends in Gestural Control of Music*, 41–70.

Bongers, B. (2006). *Interactivation: Towards an e-cology of people, our technological environment, and the arts.* Ph. D. thesis, Vrije Universiteit Amsterdam.

Brent, W. (2011). Aspects of gesture in digital musical instrument design. In *Proceedings of the International Computer Music Conference*, Huddersfield, UK.

Brown, A. and A. Sorensen (2009). Interacting with generative music through live coding. *Contemporary Music Review 28*(1), 17–29.

Brown, E. and P. Cairns (2004). A grounded investigation of game immersion. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Vienna, pp. 1297–1300.

Bryan-Kinns, N. (2012). Mutual engagement and collocation with shared representations. *International Journal of Human-Computer Studies (in press).*

Bryan-Kinns, N., R. Fencott, O. Metatla, S. Nabavian, and J. G. Sheridan (2010). Interactional sound and music: Listening to CSCW, sonification, and sound art. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, Sydney, Australia.

Bryan-Kinns, N. and F. Hamilton (2009). Identifying mutual engagement. *Behaviour & Information Technology 31*(2), 101–125.

Bryan-Kinns, N., P. G. T. Healey, and J. Leach (2007). Exploring mutual engagement in creative collaborations. In *Proceedings of the ACM Conference on Creativity and Cognition*, Washington DC, pp. 223–232.

Bryan-Kinns, N., P. G. T. Healey, J. Leach, and A. Brooker (2006). Mutual engagement in collaboration. In *Proceedings of the About-Face Workshop, Human Factors in Computing Systems*, Montreal, Canada.

Buchla, D. (2005). A history of Buchla's musical instruments. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, Vancouver, Canada.

Cadoz, C. (2009). Supra-instrumental interactions and gestures. *Journal of New Music Research 38*(3), 215–230.

Cadoz, C. and M. M. Wanderley (2000). Gesture-music. In M. M. Wanderley and M. Battier (Eds.), *Trends in Gestural Control of Music*, pp. 71–94. Paris: IRCAM.

Cage, J. (1973). *Silence: Lectures and writings.* Middletown, CT: Wesleyan University Press.

Camurri, A., P. Coletta, A. Massari, B. Mazzarino, M. Peri, M. Ricchetti, A. Ricci, and G. Volpe (2004). Toward real-time multimodal processing: EyesWeb 4.0. In *Proceedings of the AISB 2004 Convention: Motion, Emotion and Cognition*, Leeds, UK, pp. 22–26.

Camurri, A., P. Coletta, M. Ricchetti, and G. Volpe (2000). Expressiveness and physicality in interaction. *Journal of New Music Research 29*(3), 187–198.

Camurri, A., I. Lagerlöf, and G. Volpe (2003). Recognizing emotion from dance movement: Comparison of spectator recognition and automated techniques. *International Journal of Human-Computer Studies 59*(1-2), 213–225.

Camurri, A., B. Mazzarino, and G. Volpe (2004). Expressive gestural control of sound and visual output in multimodal interactive systems. In *Proceedings of the International Conference Sound and Music Computing.*

Candy, L. and E. Edmonds (2002). *Explorations in Art and Technology.* London: Springer.

Candy, L. and E. Edmonds (Eds.) (2011). *Interacting: Art, Research and the Creative Practitioner.* Faringdon, UK: Libri Publishing.

Caramiaux, B., P. Susini, T. Bianco, F. Bevilacqua, O. Houix, N. Schnell, and N. Misdariis (2011). Gestural embodiment of environmental sounds: An experimental study. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, Oslo, Norway, pp. 144–148.

Carroll, J. M. and C. Carrithers (1984). Training wheels in a user interface. *Communications of the ACM 27*(8), 800–806.

Cascone, K. (2003). Grain, sequence, system: Three levels of reception in the performance of laptop music. *Contemporary Music Review 22*(4), 101–104.

Castellano, G., R. Bresin, A. Camurri, and G. Volpe (2007a). Expressive control of music and visual media by full-body movement. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pp. 390–391.

Castellano, G., R. Bresin, A. Camurri, and G. Volpe (2007b). User-centered control of audio and visual expressive feedback by full-body movements. In A. Paiva, R. Prada, and R. Picard (Eds.), *Affective Computing and Intelligent Interaction*, Volume 4738 of *Lecture Notes in Computer Science*, pp. 501–510. Berlin/Heidelberg: Springer.

Chadabe, J. (1984). Interactive composing: An overview. *Computer Music Journal 8*(1), 22–27.

Chadabe, J. (1985, 20 October). Intelligent music business plan. Cited by Chadabe (1997, p. 332).

Chadabe, J. (1997). *Electric sound: the past and promise of electronic music.* Upper Saddle River, NJ: Prentice Hall.

Chadabe, J. (2002). The limitations of mapping as a structural descriptive in electronic instruments. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, Dublin, Ireland.

Chadabe, J. (2004). Electronic music and life. *Organised Sound 9*(1), 3–6.

Chadabe, J. (2005). The meaning of interaction, a public talk given at the workshop in interactive systems in performance (WISP). In *Proceedings of the 2005 HCSNet Conference*, Sydney, Australia. Cited by Drummond (2009).

Chafe, C. (1993). Tactile audio feedback. In *Proceedings of the International Computer Music Conference*, Tokyo, pp. 76–79.

Challis, B. (2011). Octonic: an accessible electronic musical instrument. *Digital Creativity 22*(1), 1–12.

Clanton, C. (1998). An interpreted demonstration of computer game design. In *CHI '98: CHI 98 conference summary on human factors in computing systems*, Los Angeles, CA, pp. 1–2.

Clanton, C. (2000). Lessons from game design. In E. Bergman (Ed.), *Information Applicances and Beyond*, Chapter 10, pp. 299–334. Waltham, MA: Academic Press.

Clark, A. (2007, April). Defining adaptive music. *Gamasutra.com*. `http://www.gamasutra.com/view/feature/1567/defining_adaptive_music.php`. Accessed 14 April 2012.

Clarke, E. (2006). Making and hearing meaning in performance. *The Nordic Journal of Aesthetics 18*(33-34), 24–48.

Colley, A., L. Banton, J. Down, and A. Pither (1992). An expert-novice comparison in musical composition. *Psychology of Music 20*(2), 124–137.

Collins, K. (2007). An introduction to the participatory and non-linear aspects of video games audio. In S. Hawkins and J. Richardson (Eds.), *Essays on Sound and Vision*, pp. 263–298. Helsinki: Helsinki University Press.

Collins, K. (2008). *Game Sound: An Introduction to the History, Theory, and Practice of Video Game Music and Sound Design*. MIT Press.

Collins, K. (2009). An introduction to procedural music in video games. *Contemporary Music Review 28*(1), 5–15.

Collins, N., A. McLean, J. Rohrhuber, and A. Ward (2004). Live coding in laptop performance. *Organised Sound 8*(3), 321–330.

Cook, P. R. (2001). Principles for designing computer music controllers. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, Seattle, WA, pp. 1–4.

Cook, P. R. (2009). Re-designing principles for computer music controllers: a case study of squeezevox maggie. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, Pittsburgh, PA, pp. 218–221.

Cornock, S. and E. Edmonds (1973). The creative process where the artist is amplified or superseded by the computer. *Leonardo 6*(1), 11–16.

Costello, B. (2007). A pleasure framework. *Leonardo 40*(4), 370–371.

Costello, B. (2009). *Play and the experience of interactive art.* Ph. D. thesis, University of New South Wales, Sydney.

Costello, B., L. Muller, S. Amitani, and E. Edmonds (2005). Understanding the experience of interactive art: Iamascope in Beta_space. In *Proceedings of the Second Australasian Conference on Interactive Entertainment*, Sydney, Australia, pp. 49–56.

Costello, B. M. and E. A. Edmonds (2009). Directed and emergent play. In *Proceedings of the ACM Conference on Creativity and Cognition*, Berkeley, California, USA, pp. 107–116.

Cox, A. (2001). The mimetic hypothesis and embodied musical meaning. *Musicae Scientiae 5*(2), 195–212. Cited by Brent (2011).

Creswell, J. W. (2008). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches* (Third ed.). London: Sage.

Cross, I. (1995). "The analysis and cognition of melodic complexity" by Eugene Narmour (book review). *Music Perception 12*(4), 486–509.

Csikszentmihalyi, M. (1975). *Beyond Boredom and Anxiety.* San Franscisco, CA: Jossey-Bass. Cited by Rettie (2001).

Csikszentmihalyi, M. (1992/2002). *Flow: The psychology of optimal experience* (Second ed.). New York: Harper & Row.

Csikszentmihalyi, M. and J. LeFevre (1989). Optimal experience in work and leisure. *Journal of Personality and Social Psychology 56*(5), 815 – 822.

Dannenberg, R. B. and J. Bates (1995). A model for interactive art. In *Proceedings of the Fifth Biennial Symposium for Arts and Technology*, Connecticut College, CT, pp. 103–111.

Davidson, J. W. (1993). Visual perception of performance manner in the movements of solo musicians. *Psychology of Music 21*(2), 103–113.

Davies, E. (2001). *Beyond Dance.* New York: Routledge.

Dayton, T. (1990). *Drama Games.* Deerfield Beach, FL: Health Communications, Inc.

deMarrais, K. (2004). Qualitative interview studies: Learning through experience. In K. deMarrais and S. D. Lapan (Eds.), *Foundations for Research: Methods of Inquiry in Education and the Social Sciences*, Chapter 4, pp. 51–68. Mahwah, NJ: Lawrence Erlbaum Associates.

Denzin, N. K. and Y. S. Lincoln (2005). Introduction. In N. K. Denzin and Y. S. Lincoln (Eds.), *The Sage Handbook of Qualitative Research* (Third ed.)., Chapter 1. London: Sage.

Deweppe, A., M. Lesaffre, and M. Leman (2009). Discerning user profiles and establishing usability for interactive music applications that use embodied mediation technology. In *Proceedings of the 7th Triennial Conference of the European Society for the Cognitive Sciences of Music Conference*, Jyväskylä, Finland.

Dipper, G. (2009). Interactive interfaces: Installations produced at the ZKM|IMA. *Organised Sound 14*(3), 286–298.

Dix, A. (2004). *Human-Computer Interaction.* Harlow, UK: Prentice hall.

Dobrian, C. and D. Koppelman (2006). The 'E' in NIME: Musical expression with new computer interfaces. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, Paris, France, pp. 277–282.

Douglas, Y. and A. Hargadon (2000). The pleasure principle: immersion, engagement, flow. In *Proceedings of the Eleventh ACM Conference on Hypertext and Hypermedia*, San Antonio, TX, pp. 153–160.

Drummond, J. (2009). Understanding interactive systems. *Organised Sound 14*(2), 124–133.

Dubnov, S. (2006). Spectral anticipations. *Computer Music Journal 30*(2), 63–83.

Dubnov, S. (2010). *Information Dynamics and Aspects of Musical Perception*, Chapter 7, pp. 127–158. Berlin: Springer-Verlag.

Dubnov, S., G. Assayag, and A. Cont (2011). Audio oracle analysis of musical information rate. In *The Fifth IEEE International Conference on Semantic Computing*, pp. 567–571.

Dudas, R. (2010). "Comprovisation": The various facets of composed improvisation within interactive performance systems. *Leonardo Music Journal 20*, 33–39.

Durrani, M. (2009, 9 March). Magical challenge of video game music. *BBC News website*. http://news.bbc.co.uk/1/hi/technology/7932390.stm. Accessed 25 March 2012.

Edmonds, E. and L. Candy (2010). Relating theory, practice and evaluation in practitioner research. *Leonardo 43*(5), 470–476.

Edmonds, E., L. Muller, and M. Connell (2006). On creative engagement. *Visual Communication 5*(3), 307.

Eigenfeldt, A. (2009). The evolution of evolutionary software: Intelligent rhythm generation in Kinetic Engine. *Proceedings of the European Workshop on Evolutionary and Biologically Inspired Music, Sound, Art and Design*, 498–507.

Eigenharp (2010). Products of Interest. *Computer Music Journal 34*(1), 118–128.

Eisenberg, J. and W. F. Thompson (2003). A matter of taste: Evaluating improvised music. *Creativity Research Journal 15*(2-3), 287–296.

Else, L. (2010, 25 July). Seeing stars at the secret garden party. *New Scientist CultureLab blog*. http://www.newscientist.com/blogs/culturelab/2011/07/seeing-stars-at-the-secret-garden-party.html. Accessed 14 April 2012.

Fels, S. (2000). Intimacy and embodiment: Implications for art and technology. In *Proceedings of the ACM Workshops on Multimedia*, pp. 13–16.

Fels, S. (2004). Designing for intimacy: Creating new interfaces for musical expression. *Proceedings of the IEEE 92*(4), 672–685.

Fels, S., A. Gadd, and A. Mulder (2002). Mapping transparency through metaphor:towards more expressive musical instruments. *Organised Sound 7*(2), 109–126.

Fels, S. and M. Lyons (2011). NIME primer: A gentle introduction to creating new interfaces for musical expression. Tutorial at the *International Conference for New Interfaces for Musical Expression*, Oslo, Norway.

Fels, S. and K. Mase (1999). Iamascope: A graphical musical instrument. *Computers and Graphics 23*(2), 277–286.

Fencott, R. and N. Bryan-Kinns (2010). Hey man, you're invading my personal space! Privacy and awareness in collaborative music. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, Sydney, Australia.

Fiebrink, R., D. Trueman, C. Britt, M. Nagai, K. Kaczmarek, M. Early, M. Daniel, A. Hege, and P. Cook (2010). Toward understanding human-computer interactions in composing the instrument. In *Proceedings of the International Computer Music Conference*.

Forgas, J. P. (1995). Mood and judgment: The affect infusion model (AIM). *Psychological Bulletin 117*(1), 39.

Foucault, M. (1969). *The Archaeology of Knowledge*. London: Tavistock. Cited by Banister et al. (1994).

Fyans, A. and M. Gurevich (2011). Perceptions of skill in performances with acoustic and electronic instruments. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, Oslo, Norway.

Fyans, A., M. Gurevich, and P. Stapleton (2009a). Where did it all go wrong? a model of error from the spectators perspective. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, Pittsburg, PA.

Fyans, A., M. Gurevich, and P. Stapleton (2010). Examining the spectator experience. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, Sydney, Australia, pp. 451–454.

Fyans, A. C., M. Gurevich, and P. Stapleton (2009b). Spectator understanding of error in performance. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (extended abstracts)*, Boston, MA, pp. 3955–3960.

Gadd, A. and S. Fels (2002). Metamuse: Metaphors for expressive instruments. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, Dublin, Ireland.

Gaver, W. W., A. Boucher, S. Pennington, and B. Walker (2004, September). Cultural probes and the value of uncertainty. *Interactions 11*(5), 53–56.

Gaye, L., A. Cavaco, A. Galani, R. Clarke, J. Allen, and A. Tanaka (2010). Engagement at Culture Lab Newcastle: The SiDE project and beyond. In *Digital Futures*, Nottingham, UK.

Gelineck, S. and S. Serafin (2010). A practical approach towards an exploratory framework for physical modeling. *Computer Music Journal 34*(2), 51–65.

Geroch, M. S. (2004). Motion capture for the rest of us. *Journal of Computing Sciences in Colleges 19*(3), 157–164.

Ghani, J. A. and S. P. Deshpande (1994). Task characteristics and the experience of optimal flow in human-computer interaction. *Journal of Psychology 128*(4), 381.

Godlovitch, S. (1998). *Musical Performance: A Philosophical Study*. London: Routledge. Cited by Auslander (2005).

Gonzales, A. L., T. Finley, and S. P. Duncan (2009). (Perceived) interactivity: does interactivity increase enjoyment and creative identity in artistic spaces? In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Boston, MA, pp. 415–418.

Grewe, O., B. Katzur, R. Kopiez, and E. Altenmüller (2011). Chills in different sensory domains: Frisson elicited by acoustical, visual, tactile and gustatory stimuli. *Psychology of Music 39*(2), 220–239.

Grossman, T., G. Fitzmaurice, and R. Attar (2009). A survey of software learnability: Metrics, methodologies and guidelines. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Boston, MA, pp. 649–658.

Gurevich, M. and A. C. Fyans (2011). Digital musical interactions: Performer-system relationships and their perception by spectators. *Organised Sound 16*(2), 166–175.

Gurevich, M., P. Stapleton, and P. Bennett (2009). Designing for style in new musical interactions. In *Proceedings of the conference on New Interfaces for Musical Expression (NIME09)*, Pittsburg, PA.

Gurevich, M., P. Stapleton, and A. Marquez-Borbon (2010). Style and constraint in electronic musical instruments. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, Sydney, Australia.

Gurevich, M. and J. Treviño (2007). Expression and its discontents: Toward an ecology of musical creation. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, New York, NY, pp. 106–111.

Hahn, T. and C. Bahn (2002). Pikapika – the collaborative composition of an interactive sonic character. *Organised sound 7*(3), 229–238.

Hamman, M. (1999). From symbol to semiotic: Representation, signification and the composition of music interaction. *Journal of New Music Research 28*(2), 90–104.

Harmonix Music Systems (2005). *Guitar Hero* (videogame). New York, NY: MTV Games.

Harmonix Music Systems (2007). *Rock Band* (videogame). New York, NY: MTV Games.

Hazlett, R. (2008). Using biometric measurement to help develop emotionally compelling games. In K. Isbister and N. Schaffer (Eds.), *Game Usability*. Burlington, MA: Morgan Kaufmann.

Healey, P., J. Leach, and N. Bryan-Kinns (2005). Inter-play: Understanding group music improvisation as a form of everyday interaction. In *Proceedings of Less is More – Simple Computing in an Age of Complexity*, Cambridge, UK.

Hepting, D. H., D. Gerhard, and J. Rathgaber (2005). Realtime interactive multimedia performance. In *Proceedings of the ACM Interational Conference on Computer Graphics and Interactive Techniques*, Los Angeles, CA, pp. 20.

Herzig, A. and D. Longin (2002). A logic of intention with cooperation principles and with assertive speech acts as communication primitives. In *Proceedings of the First International Joint Conference on Autonomous Agents and Multiagent Systems: Part 2*, Bologna, Italy, pp. 920–927.

Holstein, J. A. and J. F. Gubrium (2005). Interpretive practice and social action. In N. K. Denzin and Y. S. Lincoln (Eds.), *The Sage Handbook of Qualitative Research* (Third ed.)., Chapter 19. London: Sage.

Höök, K., P. Sengers, and G. Andersson (2003). Sense and sensibility: Evaluation and interactive art. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 241–248. ACM.

Hoonhout, H. C. M. (2008). Let the game tester do the talking: Think aloud and interviewing to learn about the game experience. In K. Isbister and N. Schaffer (Eds.), *Game Usability*. Burlington, MA: Morgan Kaufmann.

Hsu, W. and M. Sosnick (2009). Evaluating interactive music systems: An HCI approach. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, Pittsburgh, PA.

Hunt, A. and T. Hermann (2004). The importance of interaction in sonification. In *Proceedings of ICAD 04-Tenth Meeting of the International Conference on Auditory Display*, Sydney, Australia.

Hunt, A. and M. M. Wanderley (2002). Mapping performer parameters to synthesis engines. *Organised Sound 7*(2), 97–108.

Hunt, A., M. M. Wanderley, and R. Kirk (2000). Towards a model for instrumental mapping in expert musical interaction. In *Proceedings of the International Computer Music Conference*, Berlin, Germany, pp. 209–212.

Hunt, A., M. M. Wanderley, and M. Paradis (2003). The importance of parameter mapping in electronic instrument design. *Journal of New Music Research 32*(4), 429–440.

Huron, D. B. (2006). *Sweet anticipation: Music and the psychology of expectation*. MIT Press.

Hutt, C. (1985). Exploration and play in children. In B. Sutton-Smith and R. E. Herron (Eds.), *Child's Play*. Malabar, FL: Robert E. Krieger Publishing. Cited by Costello (2009).

Inception (2010). Reality Jockey Ltd. (developers). iPhone app. Burbank, CA: Warner Bros. Entertainment. `http://inceptiontheapp.com`. Accessed 14 April 2012.

Inskip, C. (2010). *Upbeat and quirky, with a bit of a build: Communicating meaning and meeting information needs in the music industry*. Ph. D. thesis, City University London.

Iwai, T. (2005). *Electroplankton* (Nintendo DS application). Kyoto, Japan: Nintendo.

James, J., T. Ingalls, G. Qian, L. Olsen, D. Whiteley, S. Wong, and T. Rikakis (2006). Movement-based interactive dance performance. In *Proceedings of Annual ACM International Conference on Multimedia*, pp. 470–479. ACM.

Jamieson, S. (2004). Likert scales: How to (ab)use them. *Medical Education 38*(12), 1217–1218.

Jennett, C., A. Cox, P. Cairns, S. Dhoparee, A. Epps, T. Tijs, and A. Walton (2008). Measuring and defining the experience of immersion in games. *International Journal of Human-Computer Studies 66*(9), 641–661.

Jessop, E. N., P. Torpey, and B. Bloomberg (2011). Music and technology in Death and the Powers. In *Proceedings of the International Conference on New Interfaces for Musical Expression*.

Jo, K. and A. Tanaka (2009). The music one participates in. In F. Schroeder (Ed.), *Performing Technology: User Content And The New Digital Media*, Chapter 3. Newcastle upon Tyne, UK: Cambridge Scholars Publishing.

Johnson-Laird, P. N. (1988). Freedom and constraint in creativity. In R. J. Sternberg (Ed.), *The Nature of Creativity: Contemporary psychological perspectives*, pp. 202–219. Cambridge University Press.

Johnston, A., L. Candy, and E. Edmonds (2009). Designing for conversational interaction. In *Proceedings of the International Conference on New Interfaces for Musical Expression*.

Jordà, S. (2002). Improvising with computers: A personal survey (1989–2001). *Journal of New Music Research 31*(1), 1–10.

Jordà, S. (2003). Interactive music systems for everyone: Exploring visual feedback as a way for creating more intuitive, efficient and learnable instruments. In *Proceedings of the Stockholm Music Acoustics Conference*, Stockholm, Sweden.

Jordà, S. (2004). Instruments and players: Some thoughts on digital lutherie. *Journal of New Music Research 33*(3), 321–341.

Jordà, S. (2005). *Digital Lutherie: Crafting musical computers for new musics' performance and improvisation.* Ph. D. thesis, Universitat Pompeu Fabra, Barcelona, Spain.

Jordà, S., M. Kaltenbrunner, G. Geiger, and R. Bencina (2005). The reacTable*. In *Proceedings of the International Computer Music Conference (ICMC 2005)*, Barcelona, Spain, pp. 579–582.

Kalekar, P. (2004). Time series forecasting using holt-winters exponential smoothing. Technical report, Kanwal Rekhi School of Information Technology, Mumbai, India.

Kaltenbrunner, M., S. Jordà, G. Geiger, and M. Alonso (2006). The reacTable*: A collaborative musical instrument. In *Proceedings of the IEEE International Workshops on Enabling Technologies: Infrastructure for Collaborative Enterprises*, Manchester, UK, pp. 406 –411.

Kapur, A., A. Kapur, N. Virji-Babul, G. Tzanetakis, and P. Driessen (2005). Gesture-based affective computing on motion capture data. In J. Tao, T. Tan, and R. Picard (Eds.), *Affective Computing and Intelligent Interaction*, Volume 3784 of *Lecture Notes in Computer Science*, pp. 1–7. Berlin/Heidelberg: Springer.

Karlax (2010). Products of Interest. *Computer Music Journal 34*(3), 86–96.

Kennedy, M. (1996). *The Concise Oxford Dictionary of Music* (Fourth ed.). Oxford University Press.

Kiefer, C., N. Collins, and G. Fitzpatrick (2008). HCI methodology for evaluating musical controllers: A case study. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, Genova, Italy.

Kiefer, C., N. Collins, and G. Fitzpatrick (2009). Phalanger: Controlling music software with hand movement using a computer vision and machine learning approach. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, Pittsburgh, PA.

Klein, G. (2009). Site-sounds: On strategies of sound art in public space. *Organised Sound 14*(1), 101–108.

Kobori, D., K. Kagawa, M. Iida, and C. Arakawa (2006). LINE: Interactive sound and light installation. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, Paris, France, pp. 110–113.

Kruskal, W. (1988). Miracles and statistics: The casual assumption of independence. *Journal of the American Statistical Association 83*(404), 929–940.

Kvifte, T. (2011). Musical instrument user interfaces: The digital background of the analog revolution. *The International Conference on New Instruments for Musical Expression*, Oslo, Norway. Keynote talk.

Kvifte, T. and A. Jensenius (2006). Towards a coherent terminology and model of instrument description and design. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, Paris, France, pp. 220–225.

Larssen, A., T. Robertson, and J. Edwards (2006). How it feels, not just how it looks: When bodies interact with technology. In *Proceedings of the 18th Australia Conference on Computer-Human Interaction*, Sydney, Australia, pp. 329–332.

Leder, H., B. Belke, A. Oeberst, and D. Augustin (2004). A model of aesthetic appreciation and aesthetic judgments. *British Journal of Psychology 95*(4), 489–508.

Lehmann, A. C. (1997). The acquisition of expertise in music: Efficiency of deliberate practice as a moderating variable in accounting for sub-expert performance. In I. Deliège and J. A. Sloboda (Eds.), *Perception and Cognition of Music*, pp. 161–187. Hove, UK: Psychology Press.

Leong, T. W., L. Gaye, A. Tanaka, R. Taylor, and P. C. Wright (2011). The user in flux: Bringing HCI and digital arts together to interrogate shifting roles in interactive media. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (extended abstracts)*, Vancouver, Canada, pp. 45–48.

Lerdahl, F. and C. Krumhansl (2007). Modeling tonal tension. *Music Perception 24*(4), 329–366.

Levitin, D. J., S. McAdams, and R. L. Adams (2002). Control parameters for musical instruments: A foundation for new mappings of gesture to sound. *Organised Sound 7*(2), 171–189.

Lindström, E., P. Juslin, R. Bresin, and A. Williamon (2003). "Expressivity comes from within your soul": A questionnaire study of music students' perspectives on expressivity. *Research Studies in Music Education 20*(1), 23–47.

Lingel, J. and M. Naaman (2012). You should have been there, man: Live music, DIY content and online communities. *New Media & Society 14*(2), 332–349.

Linson, A. (2011). Unnecessary constraints: A challenge to some assumptions of digital musical instrument design. In *Proceedings of the International Computer Music Conference*, Huddersfield, UK, pp. 421–424.

Litman, J. and T. Jimerson (2004). The measurement of curiosity as a feeling of deprivation. *Journal of Personality Assessment 82*(2), 147–157.

Litman, J. A. (2005). Curiosity and the pleasures of learning: Wanting and liking new information. *Cognition & Emotion 19*(6), 793 – 814.

Litman, J. A. (2008). Interest and deprivation factors of epistemic curiosity. *Personality and Individual Differences 44*(7), 1585 – 1595.

Litman, J. A. (2010). Relationships between measures of I- and D-type curiosity, ambiguity tolerance, and need for closure: An initial test of the wanting-liking model of information-seeking. *Personality and Individual Differences 48*(4), 397 – 402.

Livingstone, D. and E. Miranda (2004). Composition for ubiquitous responsive sound environments. In *Proceedings of the International Computer Music Conference*, Miami, FL, pp. 321–324.

Livingstone, R. S. and W. F. Thompson (2009). The emergence of music from the Theory of Mind. *Musicae Scientiae 13*(2 suppl), 83–115.

Livingstone, S. R., R. Mühlberger, A. R. Brown, and A. Loch (2007). Controlling musical emotionality: An affective computational architecture for influencing musical emotions. *Digital Creativity 18*(1), 43–53.

Lockheart, P. (2003). A history of early microphone singing, 1925–1939: American mainstream popular singing at the advent of electronic microphone amplification. *Popular Music and Society 26*(3), 367–385.

Loewenstein, G. (1994). The psychology of curiosity: A review and reinterpretation. *Psychological Bulletin 116*(1), 75.

Lyons, M., M. Haehnel, and N. Tetsutani (2003). Designing, playing, and performing with a vision-based mouth interface. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, Montreal, Canada, pp. 116–121.

Machover, T. (2002). Instruments, interactivity, and inevitability. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, Dublin, Ireland.

Magnusson, T. (2005). ixi software: The interface as instrument. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, Vancouver, Canada, pp. 212–215.

Magnusson, T. (2009). Of epistemic tools: Musical instruments as cognitive extensions. *Organised Sound 14*(2), 168–176.

Magnusson, T. (2010). Designing constraints: Composing and performing with digital musical systems. *Computer Music Journal 34*(4), 62–73.

Magnusson, T. (2011). Confessions of a live coder. In *Proceedings of the International Computer Music Conference*, Huddersfield, UK, pp. 609–616.

Magnusson, T. and E. Hurtado (2008). The phenomenology of musical instruments: A survey. *eContact 10*(4).

Mainstone, D. (2008). Sharewear. `http://v2.nl/lab/projects/sharewear`. Accessed 14 April 2012.

Malloch, J., D. Birnbaum, E. Sinyor, and M. M. Wanderley (2006). Towards a new conceptual framework for digital musical instruments. In *Proceedings of the International Conference on Digital Audio Effects*, Montreal, Canada.

Mandryk, R. L. (2005). *Modeling user emotion in interactive play environments: A fuzzy physiological approach.* Ph. D. thesis, Simon Fraser University.

Marquez-Borbon, A., M. Gurevich, A. Fyans, and P. Stapleton (2011). Designing digital musical interactions in experimental contexts. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, Oslo, Norway, pp. 373–376.

Marshall, J., D. Airantzis, A. Angus, N. Bryan-Kinns, R. Fencott, G. Lane, F. Lesage, K. Martin, G. Roussos, J. Taylor, L. Warren, and O. Woods (2010). Sensory threads. *Leonardo 43*(2), 196–197.

Marshall, M. T. and M. M. Wanderley (2006). Vibrotactile feedback in digital musical instruments. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pp. 226–229.

McAlpine, K. B., M. Bett, and J. Scanlan (2009). Approaches to creating real-time adaptive music in interactive entertainment: A musical perspective. In *Proceedings of the 35th AES International Conference*, London, UK.

McPherson, A. (2010). The magnetic resonator piano: Electronic augmentation of an acoustic grand piano. *Journal of New Music Research 39*(3), 189 – 202.

McQuaid, H. L., A. Goel, and M. McManus (2003). When you can't talk to customers: using storyboards and narratives to elicit empathy for users. In *Proceedings of the 2003 international conference on Designing pleasurable products and interfaces*, Pittsburgh, PA, pp. 120–125.

Melo, A. and G. A. Wiggins (2003). A connectionist approach to driving chord progressions using tension. In *AISB Symposium on Creativity in Arts and Science*, Aberystwyth, UK.

Meyer, L. (1961). *Emotion and meaning in music.* University of Chicago Press.

Michelsen, C. D., W. D. Dominick, and J. E. Urban (1980). A methodology for the objective evaluation of the user/system interfaces of the MADAM system using software engineering principles. In *Proceedings of the ACM Southeast Regional Conference*, pp. 103–109. Cited by Grossman et al. (2009).

Miller, R. B. (1968). Response time in man-computer conversational transactions. In *Proceedings of the AFIPS Fall Joint Computer Conference, Part I*, San Francisco, CA, pp. 267–277.

Minsky, M. (1981). Music, mind, and meaning. *Computer Music Journal 5*(3), 28–44.

Mion, L., G. D'Incà, A. Götzen, and E. Rapanà (2010). Modeling expression with perceptual audio features to enhance user interaction. *Computer Music Journal 34*(1), 65–79.

Miranda, E. R. and M. M. Wanderley (2006). *New digital musical instruments: control and interaction beyond the keyboard.* Middleton, WI: AR Editions.

Moog Piano Bar (2005). Products of interest. *Computer Music Journal 29*(1), 104–113.

Moore, F. R. (1988). The dysfunctions of MIDI. *Computer Music Journal 12*(1), 19–28.

Moore, J., D. Wegner, and P. Haggard (2009). Modulating the sense of agency with external cues. *Consciousness and Cognition 18*(4), 1056–1064.

Moorer, J. A. (1972). Music and computer composition. *Communications of the ACM 15*(2), 104–113.

Morales-Manzanares, R. and E. Morales (1997). Music composition, improvisation, and performance through body movements. In *AIMI International Workshop KANSEI: The Technology of Emotions*, Genoa, Italy.

Morales-Manzanares, R., E. F. Morales, R. Dannenberg, and J. Berger (2001). SICIB: An interactive music composition system using body movements. *Computer Music Journal 25*(2), 25–36.

Mori, M. (1970). The uncanny valley. *Energy 7*(4), 33–35. Japanese article cited by Mori et al. (2012).

Mori, M., K. MacDorman, and N. Kageki (2012). The uncanny valley. *Robotics Automation Magazine, IEEE 19*(2), 98 –100. English translation of the original article in Japanese by Mori (1970).

Morrison, A. J., P. Mitchell, and M. Brereton (2007). The lens of ludic engagement: Evaluating participation in interactive art installations. In *Proceedings of Annual ACM International Conference on Multimedia*, Augsburg, Germany, pp. 509–512.

Murphy, C., D. Chertoff, M. Guerrero, and K. Moffitt (2011). Creating flow, motivation, and fun in learning games. In *The Design of Learning Games*, Chapter 5. Berlin/Heidelberg: Springer-Verlag.

Murray-Browne, T. and C. Fox (2009). Global expectation-violation as fitness function in evolutionary composition. In *Proceedings of the European Workshop on Evolutionary and Biologically Inspired Music, Sound, Art and Design*, Tübingen, Germany, pp. 538–546.

Murray-Browne, T., D. Mainstone, N. Bryan-Kinns, and M. D. Plumbley (2010). The Serendiptichord: A wearable instrument for contemporary dance performance. In *Proceedings of the 128th Convention of the Audio Engineering Society*, London, UK.

Murray-Browne, T., D. Mainstone, N. Bryan-Kinns, and M. D. Plumbley (2011). The medium is the message: Composing instruments and performing mappings. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, Oslo, Norway, pp. 56–59.

Murray-Browne, T., D. Mainstone, N. Bryan-Kinns, and M. D. Plumbley (2013). The Serendiptichord: Reflections on the collaborative design process between artist and researcher. *Leonardo 46*(1), 86–87.

Murray-Rust, D. and A. Smaill (2011). Towards a model of musical interaction and communication. *Artificial Intelligence 175*(9–10), 1697 – 1721.

Nachmanovitch, S. (1990). *Free Play: The Power of Improvisation in Life and the Arts*. New York, NY: G. P. Putnam's Sons.

Nakamura, J. and M. Csikszentmihalyi (2002). The concept of flow. In C. R. Snyder and S. J. Lopez (Eds.), *Handbook of Positive Psychology*, pp. 89–105. Oxford University Press.

Narmour, E. (1990). *The analysis and cognition of basic melodic structures: The implication-realization model*. University of Chicago Press.

Narmour, E. (1992). *The analysis and cognition of melodic complexity: The implication-realization model*. University of Chicago Press.

Newton-Dunn, H., H. Nakano, and J. Gibson (2003). Block Jam: A tangible interface for interactive music. *Journal of New Music Research 32*(4), 383–393.

Ng, K. C. (2002a). Interactive gesture music performance interface. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, Dublin, Ireland.

Ng, K. C. (2002b). Sensing and mapping for interactive performance. *Organised Sound 7*(2), 191–200.

Ng, K. C. (2004). Music via motion: Transdomain mapping of motion and sound for interactive performances. *Proceedings of the IEEE 92*(4), 645–655.

Nicolls, S. (2010). Seeking out the spaces between: Using improvisation in collaborative composition with interactive technology. *Leonardo Music Journal 20*, 47–55.

Nordgren, A. (2009). Epilogue: Reactive music and invisible interfaces. In D. Mark (Ed.), *iPhone User Interface Design Projects*, pp. 235–238. New York, NY: Apress.

Norman, D. A. (2002). *The Design of Everyday Things* (Second ed.). New York, NY: Basic Books.

Nutt, C. (2008, 16 September). AGDC: A generative adaptive music system for MMO games. *Gamasutra.com*. `http://gamasutra.com/php-bin/news_index.php?story=20249`. Accessed 14 April 2012.

O'Modhrain, S. (2011). A framework for the evaluation of digital musical instruments. *Computer Music Journal 35*(1), 28–42.

O'Neill, S. (1999). Flow theory and the development of musical performance skills. *Bulletin of the Council for Research in Music Education 141*, 129–134.

OpenNI Organization (2010). *OpenNI User Guide*. `http://www.openni.org/documentation`. Accessed 5 March 2012.

Oudeyer, P.-Y., F. Kaplan, and V. Hafner (2007). Intrinsic motivation systems for autonomous mental development. *IEEE Transactions on Evolutionary Computation 11*(2), 265 –286.

Pachet, F. (2003). The Continuator: Musical interaction with style. *Journal of New Music Research 32*(3), 333–341.

Paine, G. (2002). Interactivity, where to from here? *Organised Sound 7*(3), 295–304.

Paradiso, J. (1997, December). Electronic music: New ways to play. *IEEE Spectrum 34*(12), 18–30.

Paradiso, J. A. (1999). The brain opera technology: New instruments and gestural sensors for musical interaction and performance. *Journal of New Music Research 28*(2), 130–149.

Paradiso, J. A. and S. O'Modhrain (2003). Current trends in electronic music interfaces. *Journal of New Music Research 32*(4), 345–349.

Park, C., P. H. Chou, and Y. Sun (2006). A wearable wireless sensor platform for interactive art performance. In *Proceedings of the Fourth Annual IEEE International Conference on Pervasive Computing and Communications*, pp. 13–17.

Pearce, J., M. Ainley, and S. Howard (2005). The ebb and flow of online learning. *Computers in Human Behavior 21*(5), 745–771.

Pearce, M. T. and G. A. Wiggins (2002). Aspects of a cognitive theory of creativity in musical composition. In *Proceedings of the ECAI 2002 Workshop on Creative Systems*, Lyon, France, pp. 17–24.

Pearce, M. T. and G. A. Wiggins (2004). Rethinking Gestalt influences on melodic expectancy. In *Proceedings of the 8th International Conference of Music Perception and Cognition*, Evanston, IL, pp. 367–371.

Penny, S. (1996). From A to D and back again: The emerging aesthetics of interactive art. *Leonardo Electronic Almanac 4*(4), 4–7.

Peräkylä, A. (2005). Analyzing talk and text. In N. K. Denzin and Y. S. Lincoln (Eds.), *The Sage Handbook of Qualitative Research* (Third ed.)., Chapter 34, pp. 869–886. London: Sage.

Pichlmair, M. and F. Kayali (2007). Levels of sound: On the principles of interactivity in music video games. In *Proceedings of the Digital Games Research Association Conference: Situated play*, Tokyo, Japan, pp. 424–430.

Poepel, C. (2005). On interface expressivity: A player-based study. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, Vancouver, Canada, pp. 228–231.

Potter, J. (2004). Discourse analysis as a way of analysing naturally occurring talk. In D. Silverman (Ed.), *Qualitative Research: Theory, Method and Practice* (Second ed.)., Chapter 11, pp. 200–221. London: Sage.

Preece, J., Y. Rogers, and H. Sharp (2011). *Interaction Design: Beyond Human-Computer Interaction* (Third ed.). Hoboken, NJ: Wiley.

Preissle, J. and L. Grant (2004). Fieldwork traditions: Ethnography and participant observation. In K. deMarrais and S. D. Lapan (Eds.), *Foundations for Research: Methods of Inquiry in Education and the Social Sciences*, Chapter 10, pp. 161–180. Mahwah, NJ: Lawrence Erlbaum Associates.

PrimeSense Inc. (2011). *NITE Data Sheet.* `http://www.primesense.com/en/press-room/resources/file/3-nite-data-sheet`. Accessed 14 April 2012.

Puckette, M. (1996). Pure Data: Another integrated computer music environment. In *Proceedings of the Second Intercollege Computer Music Concerts*, Tachikawa, Japan, pp. 37–41.

Rauterberg, M. (1995). About a framework for information and information processing of learning systems. In *Proceedings of the IFIP International Working Conference on Information System Concepts*, Marburg, Germany, pp. 54–69.

Reeves, S., S. Benford, C. O'Malley, and M. Fraser (2005). Designing the spectator experience. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Portland, Oregon, USA, pp. 741–750.

Reich, S. (1988). Different trains (musical work).

Rettie, R. (2001). An exploration of flow during internet use. *Internet Research: Electronic Networking Applications and Policy 11*(2), 103–113.

Richardson, P. and Y. Kim (2011). Beyond fun and games: A framework for quantifying music skill developments from video game play. *Journal of New Music Research 40*(4), 277–291.

RjDj (2008). Reality Jockey Ltd. (developers). (iPhone app). `http://rjdj.me`. Accessed 14 April 2012.

Roads, C., J. Strawn, C. Abbott, J. Gordon, and P. Greenspun (1996). *The Computer Music Tutorial.* MIT Press.

Robertson, A. (2009). *Interactive real-time music systems.* Ph. D. thesis, Queen Mary University of London.

Robson, D. (2001). Play!: Sound toys for the non musical. In *Proceedings of the International Conference on New Interfaces for Musical Expression.*

Rokeby, D. (1986). Very Nervous System. `http://homepage.mac.com/davidrokeby/vns.html`. Accessed 14 April 2012.

Rokeby, D. (1998). The construction of experience: Interface as content. In C. Dodsworth, Jr. (Ed.), *Digital Illusion: Entertaining the Future with High Technology*, Chapter 2, pp. 27–48. New York, NY: ACM Press.

Rokeby, D. (2011). Adventures in phy-gital space. *The International Conference on New Instruments for Musical Expression*, Oslo, Norway. Keynote talk.

Rollings, A. and E. Adams (2003). *on Game Design.* Berkeley, CA: New Riders Publishing.

Rosenstock, J. P. (2010). Free play meets gameplay: iGotBand, a video game for improvisers. *Leonardo Music Journal 20*, 11–12.

Rossman, G. and S. Rallis (1998). *Learning in the Field: An Introduction to Social Research.* London: Sage.

Rowe, R. (1993). *Interactive Music Systems.* MIT Press.

Rowe, R. (1999). The aesthetics of interactive music systems. *Contemporary Music Review 18*(3), 83–87.

Rutherford, J. and G. A. Wiggins (2002). An experiment in the automatic creation of music which has specific emotional content. In *Proceedings of the International Conference on Music Perception and Cognition*, Sydney, Australia, pp. 35–40.

Saint, D. and W. Stallwood (2008). *Auditorium* (videogame). Online Flash application. `http://www.playauditorium.com`. Accessed 24 March 2011.

Sales Dias, M., R. Bastos, J. Fernandes, J. Taveres, and P. Santos (2009). Using hand gesture and speech in a multimodal augmented reality environment. In *Gesture-Based Human-Computer Interaction and Simulation: Proceedings of the 7th International Gesture Workshop*, pp. 175–180.

Saltz, D. Z. (2001). Live media: Interactive technology and theatre. *Theatre Topics 11*(2), 107–130.

Sanghvi, J., G. Castellano, I. Leite, A. Pereira, P. W. McOwan, and A. Paiva (2011). Automatic analysis of affective postures and body motion to detect engagement with a game companion. In *Proceedings of the International Conference on Human-Robot Interaction*, New York, NY, pp. 305–312.

Sawyer, R. K. (2006). Group creativity: Musical performance and collaboration. *Psychology of Music 34*(2), 148–165.

Schacher, J. C. (2009). Action and perception in interactive sound installations: An ecological approach. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, Pittsburg, PA, pp. 286–289.

Schedel, M. (2002). Anticipating interactivity: Henry Cowell and the Rhythmicon. *Organised Sound 7*(3), 247–254.

Schellenberg, E. G. (1997). Simplifying the implication-realization model of melodic expectancy. *Music Perception: An Interdisciplinary Journal 14*(3), 295–318.

Schloss, W. A. (2003). Using contemporary technology in live performance: The dilemma of the performer. *Journal of New Music Research 32*(3), 239–242.

Schloss, W. A. and D. A. Jaffe (1993). Intelligent musical instruments: The future of musical performance or the demise of the performer? *Journal of New Music Research 22*(3), 183–193.

Schmidhuber, J. (2009). Driven by compression progress: A simple principle explains essential aspects of subjective beauty, novelty, surprise, interestingness, attention, curiosity, creativity, art, science, music, jokes. In G. Pezzulo, M. Butz, O. Sigaud, and G. Baldassarre (Eds.), *Anticipatory Behavior in Adaptive Learning Systems*, Volume 5499 of *Lecture Notes in Computer Science*, pp. 48–76. Springer Berlin / Heidelberg.

Schmidhuber, J. (2010). Formal theory of creativity, fun, and intrinsic motivation (1990–2010). *IEEE Transactions on Autonomous Mental Development 2*(3), 230–247.

Schnell, N. and M. Battier (2002). Introducing composed instruments, technical and musicological implications. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, Dublin, Ireland.

Searle, J. R. (1969). *Speech acts: An essay in the philosophy of language.* Cambridge University Press.

Seddon, F. and M. Biasutti (2009). A comparison of modes of communication between members of a string quartet and a jazz sextet. *Psychology of Music 37*(4), 395–415.

Sefl, A. F. (2006). Blower. In D. E. Bush and R. Kassel (Eds.), *The Organ: An Encyclopedia*, pp. 70–71. New York, NY: Routledge.

Sheridan, J. and N. Bryan-Kinns (2008). Designing for performative tangible interaction. *International Journal of Arts and Technology 1*(3), 288–308.

Sheridan, J. G. (2006). *Digital Live Art: Mediating wittingness in playful arenas.* Ph. D. thesis, University of Lancaster.

Sheridan, J. G., A. Bayliss, and N. Bryan-Kinns (2006a). iPOI. In *Proceedings of the First International Symposium on Culture, Creativity and Interaction Design*, London, UK.

Sheridan, J. G., A. Bayliss, and N. Bryan-Kinns (2006b). iPoi: acceleration as a medium for digital live art. In *Proceedings of the 8th International Conference on Ubiquitous Computing (demonstrations)*, Irvine, CA.

Sheridan, J. G., A. Bayliss, and N. Bryan-Kinns (2007). The interior life of iPoi: Objects that entice witting transitions in performative behaviour. *International Journal of Performance Arts and Digital Media 3*, 17–36.

Sheridan, J. G., N. Bryan-Kinns, and A. Bayliss (2007). Encouraging witting participation and performance in Digital Live Art. In *Proceedings of the 21st British HCI Group Annual Conference on People and Computers*, Volume 1, Lancaster, UK, pp. 13–23.

Sherry, J. L. (2004). Flow and media enjoyment. *Communication Theory 14*(4), 328–347.

Siegel, W. (2009). Dancing the music: Interactive dance and music. In R. T. Dean (Ed.), *The Oxford Handbook of Computer Music*, Chapter 10, pp. 191–213. Oxford University Press.

Silverman, D. (1998). *Harvey Sacks: Social science and conversation analysis*. Oxford University Press.

Simon, D. (2001, June). Kalman filtering. *Embedded Systems Programming*, 72–79.

Small, C. (1998). *Musicking*. Hanover, NH: Wesleyan University Press.

Spiegel, L. (1988). Music Mouse - An Intelligent Instrument (manual). `http://retiary.org/ls/progs/mm_manual/mouse_manual.html`. Accessed 19 March 2012.

St. Clair, M. and S. Leitman (2009). PlaySoundGround: An interactive musical playground. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, Pittsburgh, PA.

Stark, A. (2011). *Musicians and machines: Bridging the semantic gap in live performance*. Ph. D. thesis, Queen Mary University of London.

Stowell, D. (2010). *Making music through real-time voice timbre analysis: Machine learning and timbral control*. Ph. D. thesis, Queen Mary University of London.

Stowell, D. and S. Dixon (2011). MIR in school? Lessons from ethnographic observation of secondary school music classes. In *Proceedings of the International Society for Music Information Retrieval Conference*, Miami, FL.

Stowell, D. and A. McLean (2011). Live music-making: a rich open task requires a rich open interface. In *BCS HCI Workshop - When Words Fail: What can Music Interaction tell us about HCI?*, Newcastle upon Tyne, UK.

Stowell, D. and M. D. Plumbley (2010). Delayed decision-making in real-time beatbox percussion classification. *Journal of New Music Research 39*(3), 203–213.

Stowell, D., M. D. Plumbley, and N. Bryan-Kinns (2008). Discourse analysis evaluation method for expressive musical interfaces. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, Genova, Italy.

Stowell, D., A. Robertson, N. Bryan-Kinns, and M. D. Plumbley (2009). Evaluation of live human-computer music-making: quantitative and qualitative approaches. *International Journal of Human-Computer Studies 67*(11), 960–975.

Suchman, L. A. (1987). *Plans and situated action.* Cambridge University Press.

Sudnow, D. (2001). *Ways of the Hand: A Rewritten Account.* MIT Press.

Sykes, J. (2006). A player-centred approach to digital game design. In J. Rutter and J. Bryce (Eds.), *Understanding Digital Games.* London: Sage.

Tanaka, A. (2004a). Malleable mobile music. In *Adjunct Proceedings of the 6th International Conference on Ubiquitous Computing.*

Tanaka, A. (2004b). Mobile music making. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, Hamamatsu, Japan, pp. 154–156.

Tanaka, A. (2006). Interaction, experience, and the future of music. In K. O'Hara and B. Brown (Eds.), *Consuming Music Together: Social and Collaborative Aspects of Music Consumption Technologies*, Volume 35 of *Computer Supported Cooperative Work*, Chapter 13, pp. 271–292. Dordrecht: Springer.

Tanaka, A. (2007). Facilitating musical creativity: In collectivity and mobility. *Leonardo Electronic Almanac 15*(5,6).

Tanaka, A. and B. Bongers (2001). Global string: A musical instrument for hybrid space. In *Proceedings of Cast01//Living in Mixed Realities*, Bonn, Germany, pp. 177–181.

Tanaka, A., N. Tokui, and A. Momeni (2005). Facilitating collective musical creativity. In *Proceedings of Annual ACM International Conference on Multimedia*, Singapore, pp. 191–198.

Theremin, L. S. (1928, 14 February). Theremin. US Patent 1,658,953.

Turing, A. M. (1950). Computing machinery and intelligence. *Mind 59*(236), 433–460.

van Aart, J., C. Bartneck, J. Hu, M. Rauterberg, and B. Salem (2010). How to behave as Alice in Wonderland–about boredom and curiosity. *Entertainment Computing 1*(3–4), 125–137.

van Dijk, T. A. (1990). Discourse & Society: A new journal for a new research focus. *Discourse & Society 1*, 5–16. Cited by Antaki et al. (2003).

Van Nort, D. (2009). Instrumental listening: Sonic gesture as design principle. *Organised Sound 14*(2), 177–187.

Van Nort, D. and M. M. Wanderley (2006). Exploring the effect of mapping trajectories on musical performance. In *Proceedings of the International Conference Sound and Music Computing*, pp. 19–24.

Van Nort, D., M. M. Wanderley, and P. Depalle (2004). On the choice of mappings based on geometric properties. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pp. 87–91.

Varèse, E. and Alcopley (1968). Edgard Varèse on music and art: A conversation between Varèse and Alcopley. *Leonardo 1*(2), 187–195.

Vawter, N. (2006). Ambient addition: How to turn urban noise into music. Master's thesis, Massachusetts Institute of Technology.

Vetter, P. and D. Wolpert (2000). Context estimation for sensorimotor control. *Journal of Neurophysiology 84*(2), 1026–1034.

Vines, B. W., C. L. Krumhansl, M. M. Wanderley, and D. J. Levitin (2006). Cross-modal interactions in the perception of musical performance. *Cognition 101*(1), 80–113.

Waisvisz, M. (1985). The Hands, a set of remote MIDI-controllers. In *Proceedings of the International Computer Music Conference*, San Francisco, CA, pp. 86–89.

Wallis, I., T. Ingalls, E. Campana, and C. Vuong (2011). The instrumental hobby and long-term engagement with HCI. In *BCS HCI Workshop - When Words Fail: What can Music Interaction tell us about HCI?*, Newcastle upon Tyne, UK.

Wanderley, M. M. (2001a). Gestural control of music. In *International Workshop Human Supervision and Control in Engineering and Music*, Kassel, Germany.

Wanderley, M. M. (2001b). *Interaction musicien-instrument: Application au contrôle gestuel de la synthèse sonore*. Ph. D. thesis, L'Université Paris 6.

Wanderley, M. M. and P. Depalle (2004). Gestural control of sound synthesis. *Proceedings of the IEEE 92*(4), 632–644.

Wanderley, M. M. and N. Orio (2002). Evaluation of input devices for musical expression: Borrowing tools from HCI. *Computer Music Journal 26*(3), 62–76.

Wanderley, M. M., N. Schnell, and J. Rovan (1998). ESCHER - modeling and performing composed instruments in real-time. In *IEEE International Conference on Systems, Man, and Cybernetics*, Volume 2, San Diego, pp. 1080–1084.

Wanderley, M. M., B. W. Vines, N. Middleton, C. McKay, and W. Hatch (2005). The musical significance of clarinetists' ancillary gestures: An exploration of the field. *Journal of New Music Research 34*(1), 97 – 113.

Wang, G. (2009). Designing Smule's iPhone ocarina. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, Pittsburgh, PA.

Weir, W. (2012, 7 February). Why is it so hard for new musical instruments to catch on? *The Atlantic*. `http://www.theatlantic.com/entertainment/archive/2012/02/why-is-it-so-hard-for-new-musical-instruments-to-catch-on/252668`. Accessed 14 April 2012.

Wellcome Trust (2010). Public engagement art awards: Evaluation guidelines for applicants. `http://www.wellcome.ac.uk/Funding/Public-engagement/Application-information`. Accessed 14 April 2012.

Wessel, D. and M. Wright (2002). Problems and prospects for intimate musical control of computers. *Computer Music Journal 26*(3), 11–22.

Widmer, G., D. Rocchesso, V. Välimäki, C. Erkut, F. Gouyon, D. Pressnitzer, H. Penttinen, P. Polotti, and G. Volpe (2007). Sound and music computing: Research trends and some key issues. *Journal of New Music Research 36*(3), 169–184.

Wiggins, G. A. (2009). Semantic gap?? Schemantic schmap!! Methodological considerations in the scientific study of music. In *11th IEEE International Symposium on Multimedia*, San Diego, CA, pp. 477–482.

Wiggins, G. A., M. T. Pearce, and D. Müllensiefen (2009). Computational modelling of music cognition and musical creativity. In *Oxford Handbook of Computer Music and Digital Sound Culture*, Chapter 1. Oxford University Press.

Wilson, D. (2006, June/July). The black art of smoothing. *Electrical & Automation Technology*, 35–36.

Windsor, W. L. (2011). Gestures in music-making: Action, information and perception. In A. Gritten (Ed.), *New Perspectives on Music and Gesture*, Chapter 2, pp. 45–66. Farnham, UK: Ashgate.

Winkler, T. (1997). Creating interactive dance with the Very Nervous System. In *Proceedings of the Connecticut College Symposium on Arts and Technology*, New London, CT.

Wood, R. T., M. D. Griffiths, D. Chappell, and M. N. Davies (2004). The structural characteristics of video games: A psycho-structural analysis. *CyberPsychology & Behavior 7*(1), 1–10.

Wright, P. and J. McCarthy (2010). *Experience-Centered Design*. Morgan & Claypool Publishers.

Yin, R. K. (1994). *Case study research: design and methods*. Thousand Oaks, CA: Sage.

Zadel, M. and G. Scavone (2006). Laptop performance: Techniques, tools, and a new interface design. In *Proceedings of the International Computer Music Conference*, pp. 643–648.

Zimmerman, E. (2004). Narrative, interactivity, play and games: Four naughty concepts in need of discipline. In N. Wardrip-Fruin and P. Harrigan (Eds.), *First Person: New Media as Story, Performance, and Game*, pp. 154–164. MIT Press. Cited by Costello (2009).